
Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules*

THOMAS A. HALGREN

*Department of Molecular Design and Diversity, Merck Research Laboratories, P.O. Box 2000,
Rahway, New Jersey 07065*

Received 20 March 1995; accepted 31 August 1995

ABSTRACT

This article describes the extension of the Merck Molecular Force Field (MMFF94) to a much broader range of organic systems. It also describes a preliminary parameterization of MMFF94 for the hydronium and hydroxide ions and for various halide, alkali, and alkali earth ions as well as for such "protein" metals as Zn^{2+} , Ca^{2+} , Cu^{2+} , Cu^+ , Fe^{2+} , and Fe^{3+} . The extension employed computational data on charge distributions, molecular geometries, and conformational energies for a series of oxysulfur (particularly sulfonamide) and oxyphosphorous compounds and for a diverse set of small molecules and ions not covered in the core parameterization. It also employed experimental data for approximately 2800 good-quality structures extracted from the Cambridge Structural Database (CSD). Some of the additional computational data were used to extend the explicit parameterization of electrostatic interactions and to more widely define a useful additive approximation for the "bond polarity" parameters (bond charge increments) used in MMFF94. Both the experimental and computational data served to define reference bond lengths and angles that the extended force field uses in conjunction with force constants obtained from carefully calibrated empirical rules. The extended torsion parameters consist partly of explicit parameters derived to reproduce MP2/6-31G* conformational energies and partly of "default parameters" provided by empirical rules patterned after those used in DREIDING and UFF but calibrated, where possible,

*This article refers to Supplementary Material available from the author upon request or via the Internet at ftp.wiley.com/public/journals/jcc/suppmat/17/490 or <http://journals.wiley.com/jcc>

against computationally derived MMFF94 torsion parameters. Comparisons to experimental data show that MMFF94 reproduces crystallographic bond lengths and bond angles with relatively modest root mean square (rms) deviations of ~ 0.02 Å and 2° , respectively. © 1996 by John Wiley & Sons, Inc.

Introduction

This is the fifth and concluding article in a series introducing MMFF94, the initial published version of the Merck Molecular Force Field.^{1–4} This article describes (i) the extension of MMFF94's core parameterization to a much broader range of organic and bio-organic structures, and (ii) the preliminary parameterization of MMFF94 for a series of ions commonly encountered in bio-organic modeling.

The experimental data used in extending MMFF94's parameterization are described in the next section. The third section summarizes the MMFF94's functional form, and the fourth section defines a variety of "properties" associated with the MMFF atom types. These atom-type properties are employed in various empirical-rule and parameter-assignment procedures. Next, the fifth through tenth sections describe the use of experimental data, empirical rules, and additional computational data in assigning parameters not defined in the core portion of MMFF94.^{2–4} The eleventh section extends the parameterization in a preliminary way to "protein" metals such as Zn^{2+} , Ca^{2+} , Mg^{2+} , Cu^{2+} , Cu^+ , Fe^{2+} , and Fe^{3+} , as well as to the hydronium and hydroxide ions and to various halide, alkali, and alkali earth ions. The twelfth section then characterizes the performance of the extended force field against the crystallographic structures used in its derivation and against a smaller set of structures previously used in other assessments of molecular force fields. The final section summarizes this work and offers conclusions.

Experimental Data Used in Extending MMFF94

The experimental data used in the extension of MMFF94 were obtained from Version 5.04 (October 1992) of the Cambridge Structural Database

(CSD).⁵ We began by extracting 139 organic neutron-diffraction structures that had *R* factors of 0.075 or less and that had at most 60 atoms. To this small initial set, we added X-ray structures of similar size in groups, first asking for those having *R* factors of 0.05 or lower; to obtain sufficient data, we ultimately allowed structures having *R* factors up to 0.075. Selection screens also required, for example, that the entry be "organic," that a residue contain from 1 to 12 carbon atoms, that neither disorder nor bond discrepancies be present, and that the rms deviation for C—C bonds be 0.010 Å or better. Also, we accepted only structures containing H, C, N, O, F, Si, P, S, Cl, Br, and I. This procedure yielded 3777 CSD structures.

We prepared the extracted structures for input to OPTIMOL,¹ the host molecular-mechanics platform for MMFF, by using MAX_MOL⁶ to assign single and multiple bonds using a distance criterion. We then used a program derived from OPTIMOL to partition the resultant structures into sets according to whether or not MMFF, as then constituted, could properly deal with them. We examined all rejected structures graphically, and usually discarded those for which no hydrogen atoms were present. For some having "interesting" chemistry, however, we added missing hydrogens using the standard rules incorporated into the IDEALIZE module of AMF.⁷ We also used AMF to repair structures for which bond orders had been assigned erroneously. Other structures that could not be handled by OPTIMOL embodied chemistry not covered by the existing MMFF atom-typing procedure. We discarded some such cases—for example, those involving phosphonium or sulfonium salts or containing carbon-based or very highly delocalized anionic centers—but accommodated other by defining additional atom types and by corresponding enhancing OPTIMOL's atom-typing capabilities. Ultimately, we retained 2861 CSD structures for use in extending MMFF94's parameterization. Some replicate structures (chemically identical structures determined under differing experimental conditions) were inadvertently included, but most represent distinct chemical entities.

Form of the Merck Molecular Force Field

The MMFF94 energy expression can be written as:

$$E_{\text{MMFF}} = \sum \text{EB}_{ij} + \sum \text{EA}_{ijk} + \sum \text{EBA}_{ijk} + \sum \text{EOOP}_{ijk;l} + \sum \text{ET}_{ijkl} + \sum \text{Ev}d\text{W}_{ij} + \sum \text{EQ}_{ij} \quad (1)$$

The seven constituent terms described bond stretching, angle bending, stretch-bend, out-of-plane bending, torsion, van der Waals, and electrostatic interactions, respectively, and are defined as summarized below. Each is obtained in kilocalories per mole when distances and angles are given in angstroms and degrees, respectively¹⁻⁴:

$$\text{EB}_{ij} = 143.9325 \frac{k b_{ij}}{2} \times \Delta r_{ij}^2 (1 + cs \Delta r_{ij} + 7/12 cs^2 \Delta r_{ij}^2) \quad (2)$$

$$\text{EA}_{ijk} = 0.043844 k a_{ijk} \Delta \vartheta_{ijk}^2 (1 + cb \Delta \vartheta_{ijk}) \quad (3)$$

$$\text{EA}_{ijk} = 143.9325 \frac{k a_{ijk}}{2} (1 + \cos \vartheta_{ijk}) \quad (\text{for linear bond angles}) \quad (4)$$

$$\text{EBA}_{ijk} = 2.51210 (k b a_{ijk} \Delta r_{ij} + k b a_{kjl} \Delta r_{kj}) \Delta \vartheta_{ijk} \quad (5)$$

$$\text{EOOP}_{ijk;l} = 0.043844 \frac{k oop_{ijk;l}}{2} \chi_{ijk;l}^2 \quad (6)$$

$$\text{ET}_{ijkl} = 0.5 (V_1 (1 + \cos \Phi) + V_2 (1 - \cos 2\Phi) + V_3 (1 + \cos 3\Phi)) \quad (7)$$

$$E_{\text{vdW}ij} = \varepsilon_{ij} \left(\frac{1.07 R_{ij}^*}{R_{ij} + 0.07 R_{ij}^*} \right)^7 \times \left(\frac{1.12 R_{ij}^{*7}}{R_{ij}^7 + 0.12 R_{ij}^{*7}} - 2 \right) \quad (8)$$

$$R_{ij}^* = A_i \alpha_i^{1/4} \quad (9)$$

$$R_{ij}^* = 0.5 (R_{ii}^* + R_{jj}^*) (1 + 0.2 (1 - \exp(-12 \gamma_{ij}^2))) \quad (10)$$

$$\gamma_{ij} = (R_{ii}^* - R_{jj}^*) / (R_{ii}^* + R_{jj}^*) \quad (11)$$

$$\varepsilon_{ij} = \frac{181.16 G_i G_j \alpha_i \alpha_j}{(\alpha_i / N_i)^{1/2} + (\alpha_j / N_j)^{1/2}} \frac{1}{R_{ij}^{*6}} \quad (12)$$

$$\text{EQ}_{ij} = 332.0716 q_i q_j / (D(R_{ij} + \delta)^n), \quad (13)$$

where n normally is taken as 1.

Throughout this study, we adopt the convention that a specific atom involved in a force-field interaction is designated by i, j, k, \dots and that the corresponding MMFF numeric atom type (or sometimes, as noted, the atomic species) is designated by I, J, K, \dots . This notation makes explicit, for example, that the force constant $k b_{ij}$ and the reference bond length r_{ij}^0 for the i - j bond in $\Delta r_{ij} = r_{ij} - r_{ij}^0$ [eq. (2)] depend on the associated MMFF atom types I and J , whereas the bond distance r_{ij} depends on the atomic coordinates.

MMFF94 Atom Types and Associated Properties

The previously published set of MMFF atom types⁸ has been extended to accommodate finer distinctions needed for MMFF94's core parameterization and to cover the broader range of chemistry represented in the CSD data. A full list of the current MMFF symbolic atom types and corresponding numeric atom types is given in Table III of Part I.¹

Table I defines atomic properties associated with the MMFF numeric atom types. These quantities are used in setting up the MMFF94 energy expression and in determining values for missing parameters from defined empirical rules. The first two columns list the MMFF numeric atom type and the associated atomic number. Column 3, labeled "crd," specifies the mandatory number of bonded neighbors, and column 4, labeled "val", gives the total number of bonds made to that atom type. Where an entry shows a two-digit number, such as "34" for atom types 54 and 55, the formal valency may be either of the two cojoined values—for example, either "3" or "4." Entries of "1" in column 5 ("pilp") define those atom types that have a "pi lone pair" capable of participating in resonance interactions with, say, an adjacent multiple bond. Column 6, headed "mltb," specifies cases in which double ("2") or triple ("3") bonds are expected to be made to an atom having the listed atom type. Entries of "1" in this column designate cases in which "intermediate" bond shortenings are employed in the empirical rule for bond lengths discussed later. These entries roughly correspond to cases in which strong delocalization to a neighboring atom can be expected. Column 7 then specifies, through an entry of "1," the set of "aromatic" atom types, and column 8 ("lin") similarly delineates dicoordinate atom types J that give rise to

TABLE I.
Properties Associated With the MMFF94 Atom Types.

atype ^a	at.no. ^b	crd ^c	val ^d	pilp ^e	mltb ^f	arom ^g	linh	sbmb ⁱ	atype ^a	at.no. ^b	crd ^c	val ^d	pilp ^e	mltb ^f	arom ^g	linh	sbmb ⁱ
1	6	4	4	0	0	0	0	0	49	8	3	3	0	0	0	0	0
2	6	3	4	0	2	0	0	1	50	1	1	1	0	0	0	0	0
3	6	3	4	0	2	0	0	1	51	8	2	3	0	2	0	0	0
4	6	2	4	0	3	0	1	1	52	1	1	1	0	0	0	0	0
5	1	1	1	0	0	0	0	0	53	7	2	4	0	2	0	1	0
6	8	2	2	1	0	0	0	0	54	7	3	4	0	2	0	0	1
7	8	1	2	0	2	0	0	0	55	7	3	34	0	1	0	0	0
8	7	3	3	1	0	0	0	0	56	7	3	34	0	1	0	0	0
9	7	2	3	0	2	0	0	1	57	6	3	4	0	2	0	0	1
10	7	3	3	1	1	0	0	0	58	7	3	4	0	1	1	0	1
11	9	1	1	1	0	0	0	0	59	8	2	2	1	1	1	0	0
12	17	1	1	1	0	0	0	0	60	6	1	3	0	3	0	0	0
13	35	1	1	1	0	0	0	0	61	7	2	4	0	3	0	1	0
14	53	1	1	1	0	0	0	0	62	7	2	2	1	0	0	0	0
15	16	2	2	1	0	0	0	0	63	6	3	4	0	2	1	0	1
16	16	1	2	0	2	0	0	0	64	6	3	4	0	2	1	0	1
17	16	3	4	0	2	0	0	0	65	7	2	3	0	2	1	0	0
18	16	4	4	0	0	0	0	0	66	7	2	3	0	2	1	0	0
19	14	4	4	0	0	0	0	0	67	7	3	4	0	2	0	0	1
20	6	4	4	0	0	0	0	0	68	7	4	4	0	0	0	0	0
21	1	1	1	0	0	0	0	0	69	7	3	4	0	1	1	0	0
22	6	4	4	0	0	0	0	0	70	8	2	2	1	0	0	0	0
23	1	1	1	0	0	0	0	0	71	1	1	1	0	0	0	0	0
24	1	1	1	0	0	0	0	0	72	16	1	1	1	1	0	0	0
25	15	4	4	0	0	0	0	0	73	16	3	3	0	0	0	0	0
26	15	3	3	1	0	0	0	0	74	16	2	4	0	2	0	0	0
27	1	1	1	0	0	0	0	0	75	15	2	3	0	2	0	0	1
28	1	1	1	0	0	0	0	0	76	7	2	2	1	0	0	0	0
29	1	1	1	0	0	0	0	0	77	17	4	4	0	0	0	0	0
30	6	3	4	0	2	0	0	1	78	6	3	4	0	2	1	0	1
31	1	1	1	0	0	0	0	0	79	7	2	3	0	2	1	0	0
32	8	1	12	1	1	0	0	0	80	6	3	4	0	2	0	0	1
33	1	1	1	0	0	0	0	0	81	7	3	4	0	1	1	0	1
34	7	4	4	0	0	0	0	0	82	7	3	4	0	1	1	0	0
35	8	1	1	1	1	0	0	0	87	26	0	0	0	0	0	0	0
36	1	1	1	0	0	0	0	0	88	26	0	0	0	0	0	0	0
37	6	3	4	0	2	1	0	1	89	9	0	0	0	0	0	0	0
38	7	2	3	0	2	1	0	0	90	17	0	0	0	0	0	0	0
39	7	3	3	1	1	1	0	1	91	35	0	0	0	0	0	0	0
40	7	3	3	1	0	0	0	0	92	3	0	0	0	0	0	0	0
41	6	3	4	0	1	0	0	0	93	11	0	0	0	0	0	0	0
42	7	1	3	0	3	0	0	0	94	19	0	0	0	0	0	0	0
43	7	3	3	1	0	0	0	0	95	30	0	0	0	0	0	0	0
44	16	2	2	1	1	1	0	0	96	20	0	0	0	0	0	0	0
45	7	3	4	0	2	0	0	0	97	29	0	0	0	0	0	0	0
46	7	2	3	0	2	0	0	0	98	29	0	0	0	0	0	0	0
47	7	1	2	0	2	0	0	0	99	12	0	0	0	0	0	0	0
48	7	2	2	0	0	0	0	0									

^aMMFF atom type.^bAtomic number.^cNumber of bonded neighbors.^dValency (see text for an explanation of entries such as "34").^eA value of 1 indicates presence of one or more pi lone pairs.^fMultiple bond designator (see text).^gA value of 1 indicates an "aromatic" atom type.ⁱA value of 1 indicates a dicoordinate atom type that as a central atom gives rise to linear bond angles.^hSingle-bond-multiple-bond designator (see text).

linear ideal bond angles $I-J-K$. An entry of "1" in the final column, headed "sbmb" for "single-bond-multiple-bond," designates atom types that can form either a multiple bond or a delocalized single bond with an sp^2 - or sp -hybridized neighbor. An example of such a single bond is the bond connecting the two central carbons of butadiene, each of which has a numeric atom type 2. These "mbsb" entries are used in assigning the "bond-type" index BT_{ij} defined in part III³; a nonstandard bond-type index of "1" is assigned whenever a single bond (formal bond order 1) is found: (a) between atoms i and j of types that are not both aromatic and for which "sbmb" entries of "1" appear in Table I; or (b) between pairs of atoms belonging to different aromatic rings (as in the case of the connecting C—C bond in biphenyl).

Extension of Parameterization for Electrostatic Interactions

In undertaking this extension of MMFF94's parameterization, we realized that the performance of the resultant force field would depend critically on the quality of the partial atomic charges used in eq. (13). We shall first consider how such charges are assigned in the extended parameterization.

The partial atomic charges q_i used in MMFF94 are constructed from initial full or fractional "formal atomic charges" (usually zero, but, e.g., $+1/3$ for guanidinium nitrogens) by adding contributions from "bond charge increments" ω_{ki} that describe the polarity of the bonds to atom i from attached atoms k . Specifically, ω_{ki} gives the contribution to the final partial charge on an atom i of atom type I accumulated from, and at the expense of, its bonded neighbor k of atom type K . Thus, MMFF94 computes q_i as:

$$q_i = q_i^0 + \sum \omega_{ki} \quad (14)$$

where $\omega_{ki} = -\omega_{ik}$.

While new vdW parameters can probably be estimated with reasonable accuracy from data for similar atom types, we doubted that sufficiently accurate bond charge increments could be obtained in this way. Accordingly, we determined a number of additional bond charge increments from suitable quantum-mechanical data as described be-

low. Because the extended parameterization increases the number of distinct types of chemical bonds recognized by MMFF94 by nearly a factor of five (i.e., from 107 to 480), a full, explicit determination of all required bond charge increments would have been prohibitively costly. Fortunately, however, we found that the core MMFF94 bond charge increments ω_{ki} could be represented with quite reasonable accuracy as $\omega_{ki} \approx \rho_i - \rho_k$, where ρ_i and ρ_k are "partial bond charge increments" that mainly reflect the relative electronegativities of atoms of types I and K . As described below, the availability of this approximation meant that we only needed to determine explicit parameters for enough additional bond charge increments to allow a partial bond charge increment to be assigned for each of the MMFF94 atom types.

The derivation of the additional parameterized bond charge increments and the determination of the partial bond charge increments is described in the following two subsections. Also addressed is a complication that arises in systems for which negative values for fractional formal atomic charges need to be assigned.

PARAMETERIZATION OF ADDITIONAL BOND CHARGE INCREMENTS AND ASSIGNMENT OF FORMAL AND PARTIAL ATOMIC CHARGES

We used GAUSSIAN 92⁹ to determine optimized HF/6-31G* geometries for the additional set of 63 small molecules and molecular ions listed in Table II. To assist in the parameterization of bond charge increments, we then used CHELPG¹⁰ as implemented in GAUSSIAN 92 to obtain electrostatic potential fit charges for these systems. This set of compounds focuses on anionic systems and on compounds containing sulfur and phosphorus, particularly in oxygenated forms.

Much as was done in the core MMFF94 parameterization, we derived the bond charge increments for 62 new types of chemical bonds encountered in this set of molecules by fitting to the *ab initio* dipole moments. To make the fit determinate, we fixed 27 of the ω_{ki} , most of which involved terminal atoms, at values consistent with the CHELPG-derived charges, and then used PROBE¹¹ to fit the remaining bond charge increments to the HF/6-31G* dipole moments (enhanced by 10%).² We also fixed all core MMFF94 bond charge increments at their previously determined values. The least-squares fit generated the 35 optimized bond-charge-increment parameters labeled "X94" in the parameter file. That file also contains 39 con-

strained or otherwise assigned¹² bond charge increments labeled "#X94".

A complication arises when an atom of a given type, such as a type 32 terminal oxygen, carries a variable partial formal charge, for example, of $-3/4$ in the phosphate ion, of $-2/3$ in HPO_4^{-2} (where three terminal oxygens share a net ionic charge of -2), of $-1/2$ in H_2PO_4^- , and of 0 in phosphoric acid. Because MMFF94 treats the P—O bond as having the same polarity (i.e., the same bond charge increment) in each of these cases, the final charges, q_i , for the terminal oxy-

gens in these systems necessarily vary to the same extent as do the initial q_i^0 values. The CHELPG calculations, in contrast, assign the terminal oxygens a net partial atomic charge of -1.17 in PO_4^{-3} and of -0.75 in H_3PO_4 . This variation amounts to 0.42 electrons and seems chemically reasonable, both in sense and in magnitude. A straightforward application of eq. (14), on the other hand, would yield a much larger variation of 0.75 electrons in the terminal oxygen charge—too large, we thought.

To resolve this impasse, we could have assigned different phosphorus or oxygen atom types (and

TABLE II.
Molecules and Ions Used to Extend the Parameterization of Electrostatic Interactions.

Conf. Index ^a	Structure	Conf. Index ^a	Structure
AN05a	Methoxide anion	PO06j	Me—OPO ₂ —Me anion, cpoc = 180°
AN06a	Methyl sulfide anion	PO07a	Phosphoric acid
AN07a	Vinyl oxide anion	PO08a	Dihydrogen phosphate anion
AN08a	Vinyl sulfide anion	PO09a	Hydrogen phosphate dianion
AN09a	Thioacetate anion	PO10a	Phosphate trianion
AN10a	Perchlorate anion	PR02a	Methylphosphine
AN11a	Tetrazole anion	PR03a	Methylchlorophosphine
AN12a	Deprotonated vinyl amine anion	PR04a	(CH ₃) ₂ C=PH
HL11a	Cyclopropyl fluoride	PS01a	Trimethylphosphine sulfide
HL12a	Cyclopropyl chloride	SI01a	Silane
HL13a	Vinyl fluoride	SI02a	Methylsilane
HL14a	Vinyl chloride	SI03a	Methylhydroxysilane
NO01a	Nitromethane	SI04a	Methylchlorosilane
NO02a	Methyl nitrite	SO05a	Me—NHSO ₂ —Me, cnsc gauche, hnso anti
NO03a	Nitrosomethane	SO05b	Me—NHSO ₂ —Me, cnsc gauche, hnso cis
NO04a	Nitroethylene	SO05j	Me—NHSO ₂ —Me, cnso = 180°
NX01a	Methyl isocyanate	SO08a	Me—C(=O)—NHSO ₂ —Me, csnc gauche
NX02a	Dimethyldiazomethane	SO08b	Me—C(=O)—NHSO ₂ —Me, csnc anti
NX03a	Methyl azide	SO08c	Me—C(=O)—NHSO ₂ —Me, hnco cis
NX04a	Acetonitrile	SO08j	Me—C(=O)—NHSO ₂ —Me, csnc = 0°
OC02a	Protonated acetone	SO09a	Methanesulfonamide, MeSONH ₂
PO01a	Methyldihydrogenphosphate, o(t) -p-o-c gauche	SO10a	Methanesulfonic acid, hosc skew
PO01b	Methyldihydrogenphosphate, o(t) -p-o-c trans	SO11a	Sulfamide, C2v
PO02a	Phosphine oxide	SO13a	Dimethyl sulfoxide
PO03a	Trimethylphosphine oxide	SO14a	(CH ₃) ₂ S(—O)NH, NH analog of dimethylsulfone
PO04a	Me—CH ₂ PO ₂ —Me anion, ccpc gauche	SO15a	Me—SO ₂ —NH(—)—Me anion
PO04j	Me—CH ₂ PO ₂ —Me anion, ccpc = 180°	SO16a	(CH ₃) ₂ C=S=O
PO05a	Me—NHPO ₂ —Me anion, cpnh gauche	SO17a	Methylsulfinate anion, CH ₃ SO ₂ ⁻¹
PO05b	Me—NHPO ₂ —Me anion, cpnh skew	SO18a	Methylthiosulfinate anion, CH ₃ S(—O)S ⁻¹
PO05j	Me—NHPO ₂ —Me anion, cnpc = 180°	SO19a	Methylsulfonate anion, CH ₃ SO ₃ ⁻²
PO06a	Me—OPO ₂ —Me anion, cpoc gauche	SO20a	Sulfate ion, SO ₄ ⁻³
		SR13a	Thioacetone

^aAs described in ref. 1, the first two letters of the conformational index define the compound class, while the following two digits specify the member number for that class. The fifth character identifies the conformation, where a, b, c, ... denote equilibrium conformers, j, k, l, ... denote conformations defined using one or more torsion angle constraints, and t (and sometimes s) denote transition state structures.

different atom types for various oxysulfur compounds, etc.) for each discrete environment. This approach would allow different P—O bond charge increments to be used, but would lead a proliferation of atom types. Instead, we decided to retain a minimal set of atom types but to allow an initially affixed formal atomic charge to be shared with neighboring atoms. To implement this approach, we rewrite eq. (14) as

$$q_i = (1 - M_i \vartheta_i) q_i^0 + \vartheta_i \sum q_K^0 + \sum \omega_{KI} \quad (15)$$

where the sums on the right-hand side run over the $M_i = \text{crd}(I)$ atoms k of MMFF atom type K directly attached to atom i . In eq. (15), q_i^0 and q_K^0 remain the formal atomic charges assigned in the primary atom-typing procedure (e.g., those cited above for the oxyphosphorus series). The sum of the first two terms can then be understood to give the "effective" formal atomic charges that would need to be used as the q_i^0 in the simpler eq. (14). (We shall use the terms "primary" and "effective" to denote this distinction.) For the oxyphosphorus series, we take $\vartheta_i = 0.5$ for the terminal oxygens of MMFF type 32. For such an oxygen, M_i is unity and the attached phosphorus carries no primary formal atomic charge. Hence, the effective formal atomic charge on a terminal oxygen is half of the primary value, or -0.375 in PO_4^{-3} , -0.333 in HPO_4^{-2} , -0.25 in $\text{H}_2\text{PO}_4^{-}$, and 0.0 in H_3PO_4 . The variation between the two limiting cases is thus reduced to 0.375 electrons, satisfactorily close to the variation of 0.42 electrons found from the CHELPG calculations. For the central phosphorus atom, no primary formal atomic charge is assigned but the 4, 3, 2, or 1 attached terminal oxygens, which bear primary formal atomic charges of -0.75 , -0.666 , -0.50 , and 0.00 , generate effective formal atomic charges on phosphorus of -1.50 , -1.00 , -0.50 , and 0.00 , respectively.

Similarly, for a carboxylate oxygen, the primary value for q_i^0 used in eq. (15) is -0.5 , but the effective [eq. (14)] value obtained using $\vartheta_{32} = 0.5$ and $M_{32} = 1$ is -0.25 . For the carboxylate carbon the primary [eq. (15)] value for q_i^0 is 0 , but the effective [eq. (14)] value is -0.5 because each of two attached terminal oxygens transfers half of its primary formal charge of -0.5 . While charge sharing is not strictly required in this last case, we found that it was needed to allow a consistent set of electronegativity-related "partial bond charge increments" to be defined for use in eq. (17) below. The full set of formal charge adjustment factors ϑ_i is given in Table III. The values chosen for the ϑ_i

result in half of each primary formal atomic charge being shifted to directly bonded neighbors for cases in which the formal atomic charge is negative; no charge sharing is currently invoked when the formal charge is positive.

We should note that MMFF94's mechanism for assigning partial atomic charges does not accurately represent charge distributions in unsymmetrical but strongly delocalized anions such as vinyl oxide ($\text{H}_2\text{C}=\text{CH}-\text{O}^-$) and vinyl sulfide ($\text{H}_2\text{C}=\text{CH}-\text{S}^-$). To properly model such resonant charge distributions, MMFF94 might need to begin with an appropriately distributed formal atomic charge or to include explicit three-body adjustment terms akin to those used by Abraham and coworkers.¹³ No comparable problem arises in symmetrical systems such as carboxylate anions, however, because MMFF94 recognizes that the two terminal oxygens are equivalent and assigns a primary fractional formal atomic charge of -0.5 to each.

EMPIRICAL RULE FOR BOND CHARGE INCREMENTS

Originally, MMFF and MM2X, its predecessor,¹ estimated missing bond charge increments using the simple formula:

$$\omega_{KI} = 0.25(\chi_K - \chi_I) \quad (16)$$

where χ_I and χ_K are the Pauling electronegativities¹⁴ for the atomic species to which the atom types I and K correspond, and 0.25 is a scale factor. During the course of this work, we found that much better estimates for parameterized MMFF94 bond charge increments could be obtained from a similar formula:

$$\omega_{KI} = \rho_I - \rho_K \quad (17)$$

where ρ_I and ρ_K now depend fully on the MMFF atom type and not just on the atomic species. Because of the additive decomposition they give of the bond charge increments ω_{KI} , we refer to the ρ_i as "partial bond charge increments." Comparison to eq. (16) indicates that in essence they reflect the effective electronegativities of the MMFF atom types.

We used an iterative procedure to derive the ρ_i from known bond charge increments ω_{KI} . To make the fit determinate, we defined saturated carbon (MMFF atom type 1) as having ρ_1 identically equal to 0 . We then calculated initial values of ρ_i for atom types bonded to atom type 1 from the

TABLE III.
MMFF94 Partial Bond Charge Increments (ρ) and Formal-Charge Adjustment Factors (ϑ).

MMFF type	ρ	ϑ	MMFF type	ρ	ϑ
1	0.000	0.00	42	-0.757	0.00
2	-0.135	0.00	43	-0.326	0.00
3	-0.095	0.00	44	-0.237	0.00
4	-0.200	0.00	45	-0.260	0.00
5	-0.023	0.00	46	-0.429	0.00
6	-0.243	0.00	47	-0.418	0.00
7	-0.687	0.00	48	-0.525	0.00
8	-0.253	0.00	49	-0.283	0.00
9	-0.306	0.00	50	0.284	0.00
10	-0.244	0.00	51	-1.046	0.00
11	-0.317	0.00	52	-0.546	0.00
12	-0.304	0.00	53	-0.048	0.00
13	-0.238	0.00	54	-0.424	0.00
14	-0.208	0.00	55	-0.476	0.00
15	-0.236	0.00	56	-0.438	0.00
16	-0.475	0.00	57	-0.105	0.00
17	-0.191	0.00	58	-0.488	0.00
18	-0.118	0.00	59	-0.337	0.00
19	0.094	0.00	60	-0.635	0.00
20	-0.019	0.00	61	-0.265	0.00
21	0.157	0.00	62	-0.125	0.25
22	-0.095	0.00	63	-0.180	0.00
23	0.193	0.00	64	-0.181	0.00
24	0.257	0.00	65	-0.475	0.00
25	0.012	0.00	66	-0.467	0.00
26	-0.142	0.00	67	-0.099 ^a	0.00
27	0.094	0.00	68	-0.135	0.00
28	0.058	0.00	69	-0.099	0.00
29	0.207	0.00	70	-0.269	0.00
30	-0.166	0.00	71	-0.071	0.00
31	0.161	0.00	72	-0.580	0.50
32	-0.732	0.50	73	-0.200	0.00
33	0.257	0.00	74	-0.301	0.00
34	-0.491	0.00	75	-0.255	0.00
35	-0.456	0.50	76	-0.568	0.25
36	-0.031	0.00	77	-0.282	0.00
37	-0.127	0.00	78	-0.168	0.00
38	-0.437	0.00	79	-0.471 ^b	0.00
39	-0.104	0.00	80	-0.144	0.00
40	-0.264	0.00	81	-0.514	0.00
41	0.052	0.00	82	-0.099 ^a	0.00

^aAssumed equal to the partial bond charge increment for atom type 69.^bAssumed equal to the average of the partial bond charge increments for atom types 65 and 66.

associated bond charge increments. When applied to the full set of 195 parameterized bond charge increments, including those developed for the extended parameterization described in the preceding subsection, the final set of ρ_i values reproduced the parameterized ω_{KI} with a rms deviation of 0.06 electrons. This rms error is one fifth the

rms value of 0.30 electrons for these bond charge increments. Much of the error is concentrated in a relative handful of highly polar bonds; for eight such cases, errors of 0.15–0.37 electrons were found. Given that explicit bond charge increments are available for these cases, the performance of eq. (17) for cases to which it will need to be

applied may be somewhat better than the cited rms error of 0.06 electrons would suggest.

To complete the extension of the parameterization for electrostatic interactions, we utilized the derived partial bond charge increments to compute explicit bond charge increments from eq. (17) for the additional types of chemical bonds encountered in the CSD data. We then stored these computed bond charge increments in the associated parameter file as parameters identified by the label "E94." New bond types encountered in subsequent MMFF94 calculations, however, would use eq. (17) directly—or eq. (16), should one or both MMFF atom types lack a defined partial bond charge increment.

Extension of the Parameterization for Bond Stretching

PARAMETERIZATION OF ADDITIONAL REFERENCE BOND LENGTHS

This parameterization closely followed the procedure used in defining reference bond lengths for the core, computationally derived portion of MMFF94.³ It was carried out in two stages. In the fit, MP2/6-31G*-optimized geometries were obtained for the 58 equilibrium conformers shown in Table II (those for which the five-character conformational index¹ ends in "a," "b," or "c") and for the 17 additional conformers listed in Table IV. Starting from reference values equivalenced to the average MP2/6-31G* bond lengths, refined values were obtained for these reference bond lengths and, as described in the next section, for the associated reference angles. At each cycle in the iterative process, force constants calculated from an empirical rule (described later) were used. As before,³ iterative determinations of the reference bond lengths and angles were interleaved. For bonds types not encountered in the core parameterization, the newly determined bond parameters were then added to the core set. The core MMFF94 and these additional MP2/6-31G* based parameters are labeled "C94" and "X94," respectively, in the distributed parameter set.¹⁵

In the second stage, the reference geometries were obtained from the CSD data. In the fit to this data, the initial reference bond lengths were set to the average observed CSD bond lengths, and empirical-rule force constants were computed and updated after each iterative cycle. Upon convergence, parameters for bonds types included neither in the "C94" nor in the "X94" parameteriza-

TABLE IV.
Additional Molecules Used to Extend the
Parameterization of Bond and
Angle Interactions.

Conf. Index ^a	Structure
PR01a	Phosphine
RY01a	Acetylene
RY02a	Propyne
SO01a	Phenyl—NHSO ₂ —phenyl, cnc gauche, hnso anti
SO01b	Phenyl—NHSO ₂ —phenyl, cnc gauche, hnso cis
SO02a	Phenyl—NHSO ₂ —methyl, cnc gauche
SO02b	Phenyl—NHSO ₂ —methyl, cnc skew, hncc cis
SO02c	Phenyl—NHSO ₂ —methyl, cnc skew, hncc gauche
SO03a	Methyl—NHSO ₂ —phenyl, cnc gauche, hnsc gauche
SO03b	Methyl—NHSO ₂ —phenyl, cnc gauche, hnsc skew
SO03c	Methyl—NHSO ₂ —phenyl, cnc anti
SO04a	Pyrrole—SO ₂ —phenyl
SO07a	Phenyl—C(=O)—NHSO ₂ —methyl, cnc gauche
SO07b	Phenyl—C(=O)—NHSO ₂ —methyl, cnc anti
SO07c	Ph—C(=O)—NHSO ₂ —methyl, o = cnh cis, cnc gauche
SO07d	Ph—C(=O)—NHSO ₂ —methyl, o = cnh cis, cnc skew
SO12a	Sulfuric acid, C2

^aSee footnote a in Table I.

tion were also added to the MMFF94 parameter set, labeled "E94" for identification. For previously encountered bond types, we chose to retain the quantum-mechanically based "C94" or "X94" values in most cases. We noted, however, 36 instances, including 6 involving core MMFF94 bond lengths,³ in which the reference value derived from the crystallographic data differed from that derived by fitting to the MP2/6-31G* structures by 0.03 Å or more. In these cases, we substituted an intermediate value representing roughly the average of the two determinations. The sets of 6 and 30 parameters modified in this way are labeled "#C94" and "#X94," respectively. In addition, we replaced the two CSD-derived reference values for bonds to hydrogen atoms with empirical-rule values; it is well known that X-ray determinations

typically yield lengths for H—X bonds that are too short by ~ 0.1 Å. These empirical-rule parameters are identified by the label “#E94” in the distributed parameter set.¹⁵

EMPIRICAL RULE FOR REFERENCE BOND LENGTHS

As noted in part I,¹ atom-type equivalencies are not invoked in assigning bond-stretching parameters. Rather, empirical rules for reference bond lengths and for bond stretching force constants are employed when no fully qualified value is available.

Reference bond lengths are assigned using a modified version of the Schomaker–Stevenson rule¹⁶ based on the form described by Blom and Haaland.¹⁷ Specifically, we take¹⁸:

$$r_{ij}^0 = r_i^0 + r_j^0 - c|\chi_i - \chi_j|^n - \delta \quad (18)$$

where r_i^0 and r_j^0 represent covalent radii for atom types i and j , χ_i and χ_j are Pauling-scale electronegativities defined by Allred and Rochow and depend only on the atomic species,¹⁹ c is a proportionality constant, n is the power dependence of the electronegativity term, and δ is a “shrinkage” factor that takes account of the fact that force-field-optimized bond lengths tend to be somewhat longer than the associated reference bond lengths. The proportionality constant, c , is taken to be 0.050 if either i or j is hydrogen, and to be 0.085 otherwise. As in Blom and Haaland’s work, the power dependence, n , is taken as 1.4. The shrinkage factor, δ , is given the value 0.008 Å. This value represents the amount by which reference bond lengths defined in the core, computationally derived portion of MMFF94 on average underlie the observed MP2/6-31G*-optimized bond lengths.

With two exceptions, covalent radii r_i^0 for single bonds are taken from Blom and Haaland, where available, or from a standard collection.²⁰ For hydrogen and for carbon, single-bond covalent radii of 0.33 and 0.77 Å (rather than 0.34 and 0.79 Å) were assigned by requiring that the empirical rule predict the core MMFF94 reference bond lengths as closely as possible. Note that these single-bond radii depend only on the atomic species and not on the full MMFF atom types. Covalent radii for multiple bonds and for resonance-delocalized bonds are assigned from the single-bond radii in the following manner. First, “hybridization indices” H_i and H_j are specified for atoms i and j . These quantities are given the value “3” (corresponding

to sp^3 hybridization) in most cases, but the value “2” if the “mltb” entry for the associated atom type in Table I is “1” or “2,” or the value “1” if the “mltb” entry is “3.” Next, a bond-order-related index, BO_{ij} , is assigned from the connectivity and the atom-type properties given in Table I. BO_{ij} is initialized to the formal i – j bond order of “1,” “2,” or “3,” but is reset to “4” if $mltb(i)$ and $mltb(j)$ are both “1,” and to “5” if one of these values is “1” and the other is “2.” Last, when i and j are aromatic atoms belonging to the same aromatic ring, BO_{ij} is set to “4” if $pilp(i)$ and $pilp(j)$ are both “0” (i.e., if neither has a pi lone pair), and otherwise to “5”; this distinction allows an appropriately smaller resonance-related shortening to be applied when $BO_{ij} = 5$ than is applied when $BO_{ij} = 4$.

The assigned H_i , H_j , and BO_{ij} values are then used as follows. First, r_i^0 and r_j^0 are set to the tabulated single-bond radii for the corresponding atomic species. Then, each is reduced by 0.04 Å if $BO_{ij} = 5$, by 0.075 Å if $BO_{ij} = 4$, by 0.17 Å if $BO_{ij} = 3$ (triple bond), and by 0.10 Å if $BO_{ij} = 2$ (full double bond). Finally, when $BO_{ij} = 1$ (single bond), the hybridization indices come into play. Specifically, r_x^0 for $x = i, j$ is reduced by 0.08 Å if $H_x = 1$ and by 0.03 Å if $H_x = 2$. The values of r_i^0 and r_j^0 which result from these adjustments are then used in eq. (18). These adjustment factors represent empirical corrections determined by examining patterns in experimentally observed and in optimized core MMFF94 bond lengths. The listed corrections for full double and triple bonds, it might be noted, are consistent with results cited by Pauling.²¹

EMPIRICAL RULE FOR BOND-STRETCHING FORCE CONSTANTS

Of the many empirical rules for bond-stretching force constants that have been proposed,²² Badger’s rule²³ is perhaps the best known. In this work, however, we use an inverse sixth-power dependence to relate the desired force constants to tabulated reference values. Such a sixth-power dependence was suggested some time ago by Roy.²⁴ More recently, Byler et al.²⁵ have shown that a similar relationship using an exponent of 5.35 accurately relates force constants and bond lengths for C–N bonds. Here for simplicity we use an exponent of six and take

$$kb_{ij} = kb_{\text{ref}}(r_{\text{ref}}^0/r_{ij}^0)^6 \quad (19)$$

where kb_{ref} and r_{ref}^0 are the force constants and associated reference bond lengths given in Table V, and r_{ij}^0 is the reference bond length retrieved from the bond-stretch parameter file or calculated from eq. (18). Note that kb_{ref} and r_{ref}^0 depend on the atomic species for i and j rather than on their full MMFF numeric atom types I and J . To simplify the notation, however, this dependence has been suppressed in writing eq. (19).

In Table V, most of the reference bond lengths r_{ref}^0 are single-bond lengths taken from the "standard geometric model" of Peterson and Csizmadia²⁶; only for the N—N, N—O, and O—S bonds do the r_{ref}^0 values differ appreciably. For the entries labeled "C94," the associated reference force constants kb_{ref} were determined by re-

quiring that the force constants kb_{ij} given by eq. (19) predict the core MMFF94 force constants about as closely as possible when the associated reference bond lengths r_{ij}^0 are used in the empirical rule. When these values of r_{ij}^0 are used, eq. (19) reproduces the full set of 107 core MMFF94 force constants with a relatively modest rms deviation of 0.75 md/Å,²⁷ about one eighth the rms value of 5.81 md/Å. The force constants kb_{ref} for the entries labeled "E94" in Table V could not be calibrated in a comparable manner because no explicitly parameterized values for kb_{ij} had been derived. Instead, these force constants were evaluated at the tabulated r_{ref}^0 bond lengths by using the Herschbach and Laurie²⁸ parameterization of Badger's rule. Should a case arise for which the

TABLE V.
MMFF94 Parameters for Empirical-Rule Force Constants for Bond Stretching Used in Eq. (19).

Atom					Atom				
<i>i</i>	<i>j</i>	r_{ref}^0 ^a	kb_{ref} ^b	Source ^c	<i>i</i>	<i>j</i>	r_{ref}^0 ^a	kb_{ref} ^b	Source ^c
H	C	1.084	5.15	C94	O	O	1.48	3.6	E94
H	N	1.001	7.35	C94	O	F	1.42	4.6	E94
H	O	0.947	9.10	C94	O	Si	1.63	5.2	E94
H	F	0.92	10.6	E94	O	P	1.66	4.7	E94
H	Si	1.48	2.3	E94	O	S	1.470	9.90	C94
H	P	1.415	2.95	C94	O	Cl	1.70	4.1	E94
H	S	1.326	4.30	C94	O	Br	1.85	3.4	E94
H	Cl	1.28	4.3	E94	O	I	2.05	1.6	E94
H	Br	1.41	4.2	E94	F	Si	1.57	6.4	E94
H	I	1.60	2.7	E94	F	P	1.54	7.1	E94
C	C	1.512	3.80	C94	F	S	1.55	6.9	E94
C	N	1.439	4.55	C94	Si	Si	2.32	1.3	E94
C	O	1.393	5.40	C94	Si	P	2.25	1.5	E94
C	F	1.353	6.20	C94	Si	S	2.15	2.0	E94
C	Si	1.86	2.6	E94	Si	Cl	2.02	3.1	E94
C	P	1.84	2.7	E94	Si	Br	2.19	2.1	E94
C	S	1.812	2.80	C94	Si	I	2.44	1.5	E94
C	Cl	1.781	2.75	C94	P	P	2.21	1.7	E94
C	Br	1.94	2.60	E94	P	S	2.10	2.4	E94
C	I	2.16	1.4	E94	P	Cl	2.03	3.0	E94
N	N	1.283	6.00	C94	P	Br	2.21	2.0	E94
N	O	1.333	5.90	C94	P	I	2.47	1.4	E94
N	F	1.36	5.9	E94	S	S	2.052	2.50	C94
N	Si	1.74	3.7	E94	S	Cl	2.04	2.9	E94
N	P	1.65	4.8	E94	S	Br	2.24	1.9	E94
N	S	1.674	3.50	C94	S	I	2.40	1.7	E94
N	Cl	1.75	3.5	E94	Cl	Cl	1.99	3.5	E94
N	Br	1.90	2.9	E94	Br	Br	2.28	2.4	E94
N	I	2.10	1.6	E94	I	I	2.67	1.6	E94

^aReference bond length (Å).

^bReference force constant (md/Å).

^cThe label "C94" denotes a parameter calibrated against core MMFF94 force constants; "E94" designates a parameter calculated from Badger's rule (see text).

parameters required for eq. (19) are not contained in the parameter file (cf. Table V), the empirical rule for kb_{ij} would employ Herschbach and Laurie's implementation of Badger's rule directly, with r_{ij}^0 given by eq. (18).

Extension of Parameterization for Angle Bending

PARAMETERIZATION OF ADDITIONAL REFERENCE ANGLES

This parameterization also took place in two stages. In the first, optimized reference angles ϑ_{ijk}^0 were obtained by fitting to the MP2/6-31G* data previously cited. In this fit, the initial values for the ϑ_{ijk}^0 were set to the average of the optimized MP2/6-31G* bond angles; the angle-bending force constants ka_{ijk} were calculated at each iterative cycle from the empirical rule to be described later. Derived parameters for angle types not encountered in the core MMFF94 parameterization were then added to the core "C94" set. These additional MP2/6-31G*-derived parameters are labeled "X94" in the distributed set.¹⁵

In the second stage, the initial values for reference angles not determined either in the core parameterization or in the fit to MP2/6-31G* data just described were set equal to the average of the observed CSD bond angles. Again, empirical-rule force constants were used and were updated after each iterative cycle. As in the previous work,³ "noncompliant" reference angles in small rings were reset to their initial values periodically. In addition, the "C94" and "X94" reference angles were reimposed after each iteration to insure that the final parameters would fit the CSD data as well as possible while remaining consistent with the *ab initio*-derived parameters. In a small number of cases, individual parameters were observed to drift considerably from their starting values without substantially affecting the force-field-optimized geometry. Prior to the final two iterative cycles, neither of which used "convergence acceleration,"³ these parameters were reset to what we judged to be physically reasonable values. Following "effective convergence," as signaled by the lack of significant further improvement in the fit to the CSD data, most newly derived parameters were added to the parameter set; these parameters are identified by the label "E94."¹⁵ However, because hydrogen atoms tend to be poorly located by X-ray data, all derived parameters involving hy-

drogens were first examined carefully. Ultimately, we retained the associated CSD-derived reference angles in 128 of 186 cases. In most of these cases, the CSD data afforded 20 or more determinations for a given angle type and displayed relatively little variation in the experimentally measured bond angle. Reference angles for the remaining 58 cases involving hydrogen atoms were assigned by analogy to values and trends found in related but better defined bond angles—such parameters are identified by the label "#94."

DEFAULT AND EMPIRICAL RULE VALUES FOR REFERENCE ANGLES

MMFF94 invokes a fairly elaborate protocol involving a series of atom-type equivalencies whenever the specific, fully qualified parameter for a given i - j - k interaction is not found in the parameter table.¹ Normally, this process will find either a specific parameter or a wildcarded "0- j -0" parameter in which the atom types I and K for peripheral atoms i and k have been successively equivalenced down to "0," which matches any starting atom type. In MMFF94, the angle-bending parameters are further qualified by an index M that represents the angle-type index AT_{ijk} .³ An extensive set of $M:0-j-0$ default parameters, generated by averaging all available "C94," "X94," and "E94" $M:I-J-K$ reference angles over I and K , is included in the parameter file.¹⁵

When even a $M:0-j-0$ default value is unavailable, a simple empirical procedure is invoked to supply a reasonable reference angle. This procedure assigns an initial value of 120° for ϑ_{ijk}^0 but substitutes a tetrahedral value of 109.45° if $\text{crd}(j) = 4$. For $\text{crd}(j) = 2$, ϑ_{ijk}^0 is set to 105° if the central atom is oxygen, to 95° if the central atom has an atomic number of greater than 10, or to 180° if the central atom is classified as linear [denoted by $\text{lin}(j) = 1$ in Table I]. If, however, $\text{crd}(j) = 3$, $\text{val}(j) = 3$, and $\text{mltb}(j) = 0$ (so that j is a tricordinate atom is not expected to be made nearly planar by strong resonance interactions), ϑ_{ijk}^0 is set to 107° if j is nitrogen and to 92° otherwise. Finally, ϑ_{ijk}^0 is set to 60° if the angle i - j - k occurs in a three-membered ring, and to 90° if i - j - k occurs in a four-membered ring.

EMPIRICAL RULE FOR ANGLE-BENDING FORCE CONSTANTS

When either a $M:0-j-0$ default value or the procedure described immediately above is used to

assign ϑ_{IJK}^0 , a further procedure must then be invoked to assign the associated angle-bending force constant. In MMFF94, the force constant (in md Å/rad²) is obtained from the previously published rule²⁹:

$$ka_{IJK} = \beta Z_I C_J Z_K (r_{IJ}^0 + r_{JK}^0)^{-1} \vartheta_{IJK}^0{}^{-2} \exp(-2D) \quad (20)$$

where Z and C are atomic parameters that depend only on the atomic number, and $D = (r_{IJ}^0 - r_{JK}^0)^2 / (r_{IJ}^0 + r_{JK}^0)^2$ is a quantity that measures the relative dissimilarity in the i - j and j - k bond lengths. The scale factor, β , normally taken to be 1.75, is reduced to 85% of this value if i , j , and k are members of a four-membered ring and to 5% of this value if these atoms form a three-membered ring (in the latter case, the three angle-bending and the three bond-stretching interactions are redundant; hence, no angle-bending force constant is strictly needed). The atomic parameters Z and C are listed in Table VI; they differ slightly from the previously published values,²⁹ because the current values have been fully optimized against the data used in the original work.

Extension of Parameterization for Stretch-Bend Interactions

The form of the stretch-bend term used in MMFF94 is given in eq. (5). The only stretch-bond parameters explicitly provided in MMFF94 are those derived from *ab initio* HF/6-31G* second derivatives, nearly all of which were obtained in its core parameterization.³ When such parameters are unavailable, default values are taken from Table VII. These values depend on the rows in the periodic table to which the constituent atoms i , j ,

and k belong; they represent averages of the explicitly derived MMFF94 stretch-bend parameters for row combinations covered in the parameterization from second derivatives or, for new row combinations, reflect trends observed in the parameterized force constants.

Extension of Parameterization for Out-of-Plane Bending at Trigonal Centers

For out-of-plane bending [eq. (6)], a request for a force constant not defined in the core parameterization is handled via a successive-equivalency protocol¹ similar to that used for angle bending. Unsatisfied requests for a specific parameter $koop_{IJK:L}$ terminate with the location of a wild-carded parameter "0-J-0:0" that depends only on the atom type J of the central atom j . For atom types covered in core MMFF94, these wildcarded parameters were generated by averaging all specifically parameterized I - J - $K:L$ force constants over I , K , and L . For newly encountered atom types (i.e., 17, 26, 43, 45, 58, 67, 73, and 82), the wild-carded default values were assigned by structural analogy to values derived for the core MMFF94 atom types.

Extension of Parameterization for Torsion Interactions

EXPLICIT PARAMETERIZATION

For this extension, we used conformational energy (CE) comparisons based on MP2/6-31G* energies at MP2/6-31G*-optimized geometries and

TABLE VI.
MMFF94 Empirical-Rule Parameters for Angle-Bending Force Constants.

Atom	Z ^a	C ^a	Atom	Z ^a	C ^a
H	1.395	—	Si	2.350 ^c	0.811
B	—	0.704	P	2.350	1.068
C	2.494 ^b	1.016	S	2.980	1.249
N	2.711	1.113	Cl	2.909	1.078
O	3.045	1.337	Br	3.017	—
F	2.847	—	As	—	0.825
			I	3.086	—

^aAtomic parameters used in eq. (20). The current version of MMFF94 does not use the parameters cited for boron and arsenic.

^bUnlike the original formulation (ref. 29), the current formulation does not treat tetrahedral carbon as a special carbon type.

^cAssumed equal to the parameter for phosphorus.

"torsion profile" (TP) comparisons based on MP2/6-31G* energies for torsionally incremented structures generated as previously described.⁴ The CE comparisons were based on 40 structures listed in Tables II, IV, and VIII for the compound series PO01, PO05, SO01, SO02, SO03, SO04, SO05, SO07, SO08, and SO09. These structures yielded 30 comparisons of relative energies. The TP comparisons covered both the foregoing and the following compound series: AN09, AN12, NO01, NO02, NO03, NO04, NX02, NX03, OC02, PO03, PO04, PO06, PO07, PO08, PO09, PR02, PR03, PR04, PS01, SI02, SI03, SI04, SO10, SO11, SO12, SO13, SO14, SO15, SO16, SO17, SO18, and SO19. A set of 493 torsion-

ally incremented structures was generated, yielding 440 comparisons of relative conformational energies. These compounds included a variety of alkyl, acyl, and aryl sulfonamides and various oxyphosphorous compounds, systems for which we felt that good quality conformational energies would especially be needed for prospective pharmaceutical and bio-organic applications of MMFF94.

As in MMFF94's core parameterization, the fitting procedure employed the program TORFIT.⁴ The conformational energies were found to depend on 98 previously unparameterized sets of V_n torsion parameters ($n = 1, 2, 3$) as well as on 20 previously parameterized sets. The latter were included in the torsion fits but were fixed at their core MMFF94 values. The former were given initial values determined from the empirical rule described later. The fit itself was carried out iteratively, as previously described.⁴ Among the 98 sets

TABLE VII.
MMFF94 Default Stretch-Bend Parameters Used in Eq. (5).

Row in Periodic Table			Force Constant	
IR	JR	KR	kba_{IJK}^a	kba_{KJI}^a
0	1	0	0.15	0.15
0	1	1	0.10	0.30
0	1	2	0.05	0.35
0	1	3	0.05	0.35
0	1	4	0.05	0.35
0	2	0	0.00	0.00
0	2	1	0.00	0.15
0	2	2	0.00	0.15
0	2	3	0.00	0.15
0	2	4	0.00	0.15
1	1	1	0.30	0.30
1	1	2	0.30	0.50
1	1	3	0.30	0.50
1	1	4	0.30	0.50
2	1	2	0.50	0.50
2	1	3	0.50	0.50
2	1	4	0.50	0.50
3	1	3	0.50	0.50
3	1	4	0.50	0.50
4	1	4	0.50	0.50
1	2	1	0.30	0.30
1	2	2	0.25	0.25
1	2	3	0.25	0.25
1	2	4	0.25	0.25
2	2	2	0.25	0.25
2	2	3	0.25	0.25
2	2	4	0.25	0.25
3	2	3	0.25	0.25
3	2	4	0.25	0.25
4	2	4	0.25	0.25

^aThe indices IR, JR, and KR represent the rows in the periodic table (with hydrogen's row denoted as "0") to which the atom types I, J, and K correspond (md Å/rad²).

TABLE VIII.
Additional Molecules Used to Extend the Parameterization of Torsion Interactions.

Conf. Index ^a	Structure
SO01j	Phenyl—NHSO ₂ —phenyl, cnsc = 180°
SO01k	Phenyl—NHSO ₂ —phenyl, cnsc gauche, nsc = 0°
SO02j	Phenyl—NHSO ₂ —methyl, cnsc = 180°
SO02k	Phenyl—NHSO ₂ —methyl, sncc = 0°
SO02s	Phenyl—NHSO ₂ —methyl, cnsc gauche, n planar
SO02t	Phenyl—NHSO ₂ —methyl, hncc cis, n planar
SO02u	Phenyl—NHSO ₂ —methyl, cnsc sk, hncc g, n planar
SO04j	Pyrrole—SO ₂ —phenyl, cnsc = 0°
SO04k	Pyrrole—SO ₂ —phenyl, nsc = 0°
SO05s	Methyl—NHSO ₂ —methyl, cnsc g, hnso a, n planar
SO05t	Methyl—NHSO ₂ —methyl, cnsc g, hnso cis, n planar
SO07j	Phenyl—C(=O)—NHSO ₂ —methyl, csnc = 0°
SO07t	Phenyl—C(=O)—NHSO ₂ —methyl, ocnh c, csnc sk, n planar
SO08t	Me—C(=O)—NHSO ₂ —me, csnc gauche, n planar
SO08t	Me—C(=O)—NHSO ₂ —me, o = cnh cis, n planar
SO09t	Methanesulfonamide, n planar
SO12a	Sulfuric acid

^aSee footnote a in Table I. The abbreviations "c" for cis, "sk" for skew, "g" for gauche, and "a" for anti are used in some of the conformational descriptions.

TABLE IX.
Comparison of MP2/6-31G* and MMFF94 Relative
Conformational Energies (kcal/mol).

Conf. A ^a	Conf. B	E(A) – E(B) ^b		Δ
		MP2	MMFF94	
PO01a	PO01b	1.31	0.89	–0.42
PO05a	PO05b	1.44	1.37	–0.07
PO05t	PO05a	1.13	0.63	–0.50
SO01b	SO01a	1.64	1.92	0.28
SO01j	SO01a	3.51	3.97	0.46
SO01k	SO01a	4.77	4.61	–0.16
SO02b	SO02a	1.17	0.57	–0.60
SO02c	SO02a	0.88	0.57	–0.32
SO02j	SO02b	1.72	2.19	0.48
SO02k	SO02a	1.52	1.05	–0.46
SO02s	SO02a	1.28	1.04	–0.24
SO02t	SO02b	2.22	0.47	–1.74
SO02u	SO02d	0.39	0.47	0.08
SO03a	SO03b	0.30	0.37	0.06
SO03c	SO03b	1.35	0.00	–1.35
SO04j	SO04a	4.73	4.76	0.04
SO04k	SO04a	1.49	1.52	0.04
SO05a	SO05b	1.74	1.17	–0.57
SO05s	SO05a	0.53	0.23	–0.30
SO05t	SO05b	2.27	1.40	–0.87
SO07c	SO07a	3.19	2.74	–0.45
SO07d	SO07a	3.30	3.90	0.60
SO07j	SO07a	5.38	4.65	–0.73
SO07t	SO07d	0.74	0.55	–0.18
SO08b	SO08a	1.55	1.47	–0.08
SO08j	SO08a	5.51	4.94	–0.57
SO08s	SO08a	0.30	0.12	–0.18
SO08t	SO08b	0.17	0.01	–0.15
SO09t	SO09b	3.16	3.56	0.40
rms conf. energy, error		2.53	0.57	

^a See Table II in ref. 1 for a description of the conformations.

^b MP2/6-31G* energies at MP2/6-31G* optimized geometries; MMFF94 energies at MMFF94-optimized geometries.

of one-, two-, and three fold torsion parameters optimized, 102 individual components ultimately were fixed at zero values, and 8 V_2 components for C=C double bonds and for bonds in aromatic rings were fixed at nonzero values chosen to be consistent with core MMFF94 values. In all, 180 nonzero V_n components were determined.

The final parameters are identified by the label "X94" in the distributed parameter set.¹⁵ These parameters reproduce 29 CE (Table IX) and 440 TP conformational comparisons with rms deviations of 0.57 and 0.73 kcal/mol, respectively, as against rms values of 2.53 and 4.90 kcal/mol. These rms deviations are higher than those of 0.31 and 0.50

kcal/mol found for CE and TP comparisons in MMFF94's core parameterization,^{1,4} possibly because the present comparisons contain a higher percentage of systems having two or more highly polar bonds in close proximity. Such systems have been found to be described less accurately,⁴ perhaps because MMFF94, like all other contemporary force fields used in general molecular modeling, too greatly simplifies the treatment of electrostatic interactions. Nevertheless, apart from a few instances in which the MMFF94 optimization failed to maintain the starting MP2/6-31G* conformation, the conformational energies are described reasonably well. As a general measure of the quality of the fit, we note that none of the 29 CE comparisons has the wrong sign for the relative conformational energy and that all but 9 of the 440 TP comparisons have the correct sign. Of the latter nine, only two involve errors in conformational energy of 0.4 kcal/mol or greater, and none is in error by more than 1 kcal/mol.

The most egregious qualitative failure occurs for conformer SO07b, where the nitrogen flattens considerably and the N—S bond rotates by $\sim 120^\circ$ during the MMFF94 optimization. This case is not reflected in the cited rms deviation, but would raise it very considerably because the SO07a structure to which SO07b collapses lies 7 kcal/mol lower on the MP2/6-31G* surface. A second qualitative failure—one that is reflected in the cited rms deviation—arises because conformer SO03c improperly optimizes to the SO03b geometry on the MMFF94 energy surface. A third occurs for SO02b and SO02c as well as for their N-planar analogs, SO02s and SO02t. For SO02b and SO02c, the two MP2/6-31G* conformers differ by just 0.3 kcal/mol on the MP2/6-31G* surface and can be interconverted by rotations of 41° about the C(ring)—N bond and of 24° about the N—S bond. Ordinarily, two equilibrium structures would not be found so close together in conformational space. Nevertheless, both are minima on the MP2/6-31G* surface, as evidenced by the fact that intermediate structures reoptimize to one or the other. The MMFF94 optimization, in contrast, yields a common (intermediate) structure from either starting point.

Torsional geometries are reasonably well described. When conformers SO03c and SO07b are excluded, the rms deviation in the MMFF94 geometries is found to be 6.53° for 1156 torsion angles in 61 equilibrium conformations. These conformers consist of those structures in Tables II, IV,

and VIII that have "a," "b," "c," or "d" as the fifth character of the conformational index¹ and are large enough to possess torsion angles. The rms deviation is only moderately higher than that found in MMFF94's core parameterization.⁴

A few years ago, we used STO-3G* data to parameterize MM2X, MMFF's predecessor force field, for a similar set of sulfonamides. When assessed against the higher quality MP2/6-31G* data used in this work, MM2X is found to perform much less accurately than MMFF94. It gives a rms deviation of 1.23 kcal/mol for the 26 CE comparisons based on SO-series conformers,³⁰ about double MMFF94's rms deviation of 0.60 kcal/mol for the same set of comparisons. MM2X, however, describes the torsional geometries about as well as MMFF94.

GENERATION OF USE OF WILDCARDED DEFAULT PARAMETERS

Requests for one-, two-, and threefold torsion parameters that cannot be resolved using the primary numeric MMFF atom types and the associated torsion-type index,⁴ TT_{IJKL} , are first handled via the "step-down" equivalency procedure outlined in part I.¹ Ultimately, a successively equivalenced request for V_1 , V_2 , and V_3 torsion parameters for an $i-j-k-l$ torsion angle may retrieve a fully wildcarded $M:0-J-K-0$ parameter (where M denotes the torsion-type index TT_{IJKL}) or, in a few cases, may access a "half-wildcarded" $M:0-J-K-L$ or $M:I-J-K-0$ parameter. If not even a wildcarded parameter is available, the empirical rule described in the next subsection is invoked.

The wildcarded parameters supplied in the parameter file were generated in one of two ways. For $J-K$ bond types covered in the "C94" (core MMFF94) or "X94" parameterizations, wildcarded parameters $M:0-J-K-0$, $M:0-J-K-L$, or $M:I-J-K-0$ were generated by averaging the explicitly determined $M:I-J-K-L$ parameters over I and L , over I , or over L , respectively. For the additional torsion interactions encountered in the extended CSD parameterization, in contrast, wildcarded $M:0-J-K-0$ parameters were generated from the empirical rule described in the following subsection and were placed in the distributed parameter file,¹⁵ identified by the label "E94" or "#E94."³¹ These parameters depend only on the torsion type and the atom types of the central two atoms.

EMPIRICAL RULE FOR TORSION PARAMETERS

The empirical rule used in MMFF94 was patterned after those used by Mayo et al. in their DREIDING force field³² and by Rappé and coworkers in UFF.³³ In MMFF94, the empirical-rule routine executes the following stepwise protocol and terminates when a stated condition is satisfied.

- (a) If either of the central atoms j and k in the $i-j-k-l$ torsion angle is designated as being linear [i.e., if $\text{lin}(J) = 1$ or $\text{lin}(K) = 1$ in Table I], zero torsion parameters are assigned.
- (b) If the atom types J and K are aromatic and the bond $j-k$ lies in an aromatic ring, a twofold potential is assigned. Independently of the value of the torsion-type index M , $V_2(M:IJKL)$, in kilocalories per mole, is given by

$$V_2 = \beta \pi_{jk} (U_j U_k)^{1/2} \quad (21)$$

where π_{jk} is related to the $j-k$ pi-bond order, and U_j and U_k are parameters that depend on the atomic species (cf. Table X). The "pi-bond order" π_{jk} is assigned as 0.5 if neither j nor k contributes a pi lone pair to the pi system [i.e., if $\text{pilp}(J)$ and $\text{pilp}(K)$, as listed in Table I, are both "0"] and otherwise as 0.3. The scale factor β is taken as 6 in most cases, but as 3 if one of the two central atoms is trivalent and the other is tetravalent, as determined from the associated $\text{val}(J)$ and $\text{val}(K)$ values.

- (c) When the $j-k$ bond has a formal bond order of 2, eq. (21) is used with $\beta = 6$ and $\pi_{jk} = 1$, provided that $\text{mltb}(J) = 2$ and $\text{mltb}(K) = 2$ (implying a full double bond). Otherwise, π_{jk} is taken as 0.4. An example of the latter

TABLE X.
MMFF94 Empirical-Rule Parameters for Torsion Interactions.

Atom	U^a	V^b	Atom	U^a	V^b
C	2.0	2.12	Si	1.25	1.22
N	2.0	1.5	P	1.25	2.40
O	2.0	0.2	S	1.25	0.49

^aUsed in eq. (21).

^bUsed in eq. (22).

case would be the formal double bond between a carbon of MMFF numeric type 57 and a nitrogen of type 55 in a resonance structure for a guanidinium cation, where $\text{mltb}(57) = 2$ but $\text{mltb}(55) = 1$. The smaller value for π_{jk} recognizes that $j-k$ is not a double bond in all resonance structures.

- (d) When both j and k are tetracoordinate [i.e., $\text{crd}(J) = 4$ and $\text{crd}(K) = 4$], a threefold potential is assigned. $V_3(M:IJKL)$ is given by:

$$V_3 = (V_J V_K)^{1/2} / N_{JK} \quad (22)$$

where V_J and V_K depend on the atomic species for j and k (cf. Table X), and N_{JK} , the number of torsion interactions about the $j-k$ bond, is obtained as the product of $\text{crd}(J) - 1$ and $\text{crd}(K) - 1$.

- (e) When atom j is tetracoordinate but k is not, a number of cases arise: (1) if k is tricoordinate but is tetravalent or is strongly delocalized [strictly, if $\text{crd}(K)$ is "3" and (i) if $\text{val}(K)$ is "4" or "34" or (ii) if $\text{mltb}(K)$ is nonzero], zero torsion parameters are returned, as this torsion bond represents a sp^3 - sp^2 rotor. (2) If k is dicoordinate but trivalent [i.e., if $\text{crd}(K) = 2$ and if either $\text{val}(K) = 3$ or $\text{mltb}(K) \neq 0$], zero torsion parameters are also returned. (3) In all other cases, eq. (22) is used and $j-k$ is treated as a saturated bond.
- (f) When atom k is tetracoordinate but j is not, the provisions of the preceding paragraph apply with the indices reversed.
- (g) The remaining cases involve central atoms j and k that, at most, are tricoordinate. When the formal $j-k$ bond order is 1 but $\text{mltb}(J)$ and $\text{mltb}(K)$ are both nonzero, or when $\text{mltb}(J)$ and $\text{pilp}(K)$ are both nonzero, or when $\text{pilp}(J)$ and $\text{mltb}(K)$ are both nonzero, $j-k$ in most cases is recognized as being a resonant interaction. A number of cases arise: (1) If both atoms are classified as having pi lone pairs [$\text{pilp}(J) = 1$ and $\text{pilp}(K) = 1$], the interaction is taken not to be resonant, and zero torsion parameters are assigned. (2) If j has a pi lone pair and k can form multiple bonds [i.e., if $\text{pilp}(J)$ and $\text{mltb}(K)$ are both nonzero], eq. (21) is used with $\beta = 6$. If j can also form strongly delocalized "single" bonds [i.e., if $\text{mltb}(J) = 1$, an example being an amide nitrogen of MMFF atom type 10], a "pi bond order" π_{jk}

of 0.5 is assigned. Otherwise, $\pi_{jk} = 0.3$ is assigned if j and k belong to carbon's row, and $\pi_{jk} = 0.15$ is assigned if either j or k does not. (3) Similarly, if k has a pi lone pair and j can form multiple bonds, case (2) applies with the indices interchanged. (4) If either j or k can form strongly delocalized "single" bonds [i.e., if either $\text{mltb}(J) = 1$ or $\text{mltb}(K) = 1$] and if either j or k is not carbon, eq. (21) is used with $\beta = 6$ and $\pi_{jk} = 0.4$. This relatively large pi bond order recognizes that the $j-k$ bond is probably a delocalized partial double bond of resonant bond order 1/2 or 1/3, an example being a C—N "single bond" in a resonance structure for a guanidinium cation [note that rule (c) above ensures that the C=N "double bond" in the same resonance structure is treated in the same way]. (5) Finally, in all other cases, $j-k$ is taken to be a "normal," but weak, interaction between multiple bonds (an example being the central C—C bond in butadiene or biphenyl). Then eq. (21) used with $\beta = 6$ and $\pi_{jk} = 0.15$.

- (h) Finally, if none of the previous conditions apply, j and k must both be saturated centers that at most are tricoordinate. In most such instances, e.g., (22) is used with the V_X parameters listed in Table X. However, if both j and k are oxygen or sulfur, each atom has two lone pairs and the interaction takes on a special nature. When j (or k) is oxygen, W_J (or W_K) is set to 2.0, but when j (or k) is sulfur, W_J (or W_K) is set to 8.0. A twofold potential $V_2(IJKL)$ is then assigned as:

$$V_2 = -(W_J W_K)^{1/2} \quad (23)$$

To place the preceding formulation of the empirical rule in context, we note that we have followed Rappé et al.³³ in allowing the U_X and V_X parameters in eqs. (21) and (22) to depend on the atomic species; Mayo et al. had suppressed such dependence in DREIDING.³² We differ from Rappé et al. in two basic ways. First, for aromatic and resonant interactions we make finer distinctions in the values assigned for twofold torsion constants. These distinctions reflect patterns found in the explicitly determined "C94" and "X94" parameters. Second, we have chosen the particular functional forms and the specific β , π_{jk} , U_X , V_X , and

W_X parameters by fitting the resultant $M:0-J-K-0$ parameters to the averaged default values obtained from the explicitly determined parameters.

When needed, the composite procedure outlined in this and the preceding subsection provides reasonable default or empirical-rule parameters. It should be kept in mind, however, that many applications will have no need for such default-level parameters, because the explicit "C94" and "X94" parameterizations cover an extensive range of chemical structure.

Extension of Parameterization for van der Waals Interactions

MMFF94 makes no use either of default parameters or of empirical rules for generating "missing" van der Waals (vdW) parameters. Atomic polarizabilities α_i needed for the additional atom types introduced in the CSD parameterization were obtained by retaining the preliminary values published earlier⁸ or were generated by analogy to "C94" (core MMFF94) parameters for similar chemical species. MMFF atom types 13, 14, 17–19, 25, 26, 42, 43, 45–48, 51, 53, 60–62, 67, 72–79, and 82 were parameterized in this way. The parameters chosen by analogy are identified by the label "X94" in the distributed parameter file.¹⁵

MMFF94 (but not MMFF93^{3,4}) also includes vdW parameters for "protein metals" such as Zn^{2+} , Ca^{2+} , Cu^{2+} , Cu^+ , and for high-spin Fe^{2+} and Fe^{3+} , as well as for F^- , Cl^- , Br^- , OH^- , Li^+ , Na^+ , Mg^{2+} , and H_3O^+ .¹⁵ This parameterization involved atom types 35, 49, and 87–99. The scale parameters A_i and G_i and the effective-electron numbers N_i were assigned by analogy to other cases, but the polarizabilities, α_i , which determined the vdW radii through eq. (9), were explicitly parameterized. The polarizability of Zn^{+2} was obtained by considering the coordination geometries for a number of ligands bound to thermolysin for which crystallographic structures were available from the protein data bank at Brookhaven National Laboratory.³⁴ Parameters for the remaining ions were then obtained by suitably adjusting the MMFF energies and structures of mono- and trihydrates for singly charged ions (Table XI) and of tri- and hexahydrates for multiply charged ions (Table XII). For most of the singly charged ions, the computed association energies could be compared with enthalpies of association determined

TABLE XI.
MMFF94 Results for Hydrates of Singly Charged Ions.

X	Ionic Radius ^b	$X(H_2O)$		$X(H_2O)_3^a$	
		$R_{X\cdots O}^c$	ΔE^d	$R_{X\cdots O}^c$	ΔE^d
F^-	1.36	2.65	−20.4	2.72 ^e	−56.8
Cl^-	1.81	3.18	−14.3	3.25 ^f	−40.9
Br^-	1.95	3.33	−13.1	3.41 ^g	−37.8
OH^-	—	2.64	−21.7	2.72 ^h	−60.8
Li^+	0.60	1.83	−32.4	1.90	−85.0
Na^+	0.95	2.21	−24.1	2.26	−65.4
K^+	1.33	2.61	−18.3	2.66	−50.4
H_3O^+	—	2.49	−28.3	2.52 ^h	−78.5
NH_4^+	—	2.59	−22.5	2.62 ^h	−62.3
Cu^+	0.96	2.23	−23.9	2.28	−64.9

^a D_3 symmetry except as noted.

^bFrom Pauling (ref. 36).

^cDistance to oxygen of water (Å); in hydronium and ammonium hydrates, from O and N atoms, respectively.

^dInteraction energy (kcal/mol); computed as energy relative to hydrate species at infinite separation.

^eUnsymmetrical; $F\cdots O$ distances are 2.67, 2.70, and 2.78 Å.

^fUnsymmetrical; $Cl\cdots O$ distances are 3.14, 3.22, and 3.39 Å.

^gUnsymmetrical; $Br\cdots O$ distances are 3.28, 3.36, and 3.58 Å.

^h C_3 symmetry.

experimentally in the gas phase.³⁵ Except for OH^- and H_3O^+ , calculated interatomic distances could also be compared to assigned ionic radii, taken from Pauling³⁶ or from Day and Selbin.³⁷ We also compared MMFF results for the monohydrates of Li^+ , Na^+ , F^- , and Cl^- to the TIP4P energies and geometries reported by Jorgensen and coworkers.³⁸ For each ionic species, we adjusted the associated polarizability α_i to make the distances R_{X-O} correlate with the ionic radius of "X" in a uniform fashion (see Tables XI and XII) and, where data were available, to bring the association energies for the hydrates roughly into line with experiment. These parameters are also labeled "X94."¹⁵

As can be seen from Tables XI and XII, the R_{X-O} distances typically lengthened by 0.05 to 0.1 Å on going from less to more highly hydrated complexes, much as might be expected. For the trihydrates, the MMFF94 distances usually exceed the listed ionic radius by about 1.40–1.45 Å for the negative ions and by about 1.30 Å for the positive ions, though the calculated R_{X-O} distances for F^- and Mg^{2+} , the smallest ions of their charge types, are about 0.05 Å shorter than would be expected

TABLE XII.
MMFF94 Results for Hydrates of Multiply Charged Ions.

X	Ionic radius ^c	X(H ₂ O) ₃ ^a		X(H ₂ O) ₆ ^b	
		$R_{X \cdots O}^d$	ΔE^e	$R_{X \cdots O}^d$	ΔE^e
Mg ²⁺	0.65	1.90	-177.3	2.08	-288.1
Ca ²⁺	0.99	2.30	-129.3	2.40	-225.6
Cu ²⁺	0.72 ^f	2.02	-162.2	2.16	-271.2
Fe ²⁺	0.76	2.06	-156.8	2.19	-264.5
Fe ³⁺	0.64	1.94	-262.8	2.06	-456.0
Zn ²⁺	0.74	2.04	-158.9	2.18	-267.1

^aD₃ symmetry.

^bO_h symmetry.

^cFrom Pauling (ref. 36).

^dDistance to oxygen of water (Å).

^eInteraction energy (kcal/mol); computed as energy relative to hydrate species at infinite separation.

^fFrom ref. 37.

from these relationships. The calculated $R_{X \cdots O}$ distances thus appear to be reasonable.

Kebarle³⁵ has reported experimentally observed gas-phase enthalpies of association for all the singly charged ions except Cu⁺. In order of their appearance in Table XI, Kebarle's values are -23.3, -13.1, -12.6, -25, -23, -22, -19.9, -24.4, and -20 kcal/mol, respectively, for the monohydrates, and -53.6, -37.5, -36.4, -56.4, -81, -60, -45.9, -69, and -45.3 kcal/mol for the trihydrates.³⁵ Comparison shows that the MMFF94 values usually are moderately more negative, as we assume should be the case for fully solvated versus gas-phase complexes.³⁹ Exceptions arise in a few cases. For F⁻, for example, the MMFF94 monohydrate energy of -20.4 kcal/mol is less negative than the experimental value of -23.3 kcal/mol, though the MMFF94 association energy for the trihydrate of -56.8 kcal/mol is more negative than experimental value of -53.6 kcal/mol. As MMFF94 is intended for use in condensed-phase simulations, we thought it more important to match the properties of the trihydrate. A second exception arises for NH₄⁺, where the MMFF94 mono- and trihydrate energies of -22.5 and -62.3 kcal/mol, respectively, are substantially more negative than the experimental values of -17.2 and -45.3 kcal/mol. We could have reduced the magnitudes of the calculated values by increasing the polarizability for quaternary nitrogen (MMFF atom type 34), thereby increasing the vdW radius and lengthening the calculated N⁺ \cdots O distances. We elected not to do so, however, because the polarizability in

question (1.00 Å³) had already been widely used in portions of the MMFF94 parameterization that had been completed before the present parameterization for ionic species was undertaken. It may be that the ammonium ion is too strongly hydrated and that its parameterization will need to be modified.

Performance of the Extended Parameterization for MMFF94

In this section we examine the ability of the extended force field to reproduce structural and energetic data obtained from the Cambridge Structural Database. The fit to the *ab initio* data on conformational energies was described previously.

FIT TO CSD DATA USED IN THE PARAMETERIZATION

As previously noted, the parameterization data consisted of 2861 sets of X-ray and neutron-diffraction structures. To assess MMFF94's ability to reproduce this structural data, we compared the MMFF94-optimized and experimental geometries for these structures. Because hydrogen atoms are not reliably placed by fits to X-ray data, we examined only bond lengths and bond angles that did not involve hydrogen atoms. In addition, because crystal packing considerations can easily distort the molecular conformation, we did not consider torsion (see, however, the following subsection). As bond angles sometimes depend significantly on the molecular conformation,⁴⁰ we applied weak penalty-function restraints to keep the torsion angles from deviating too greatly from the experimental values. These restraints added a penalty of 1.5×10^{-4} kcal/mol for each squared degree of deviation in a given torsion angle. They raise the energy by ~ 0.3 kcal/mol for a rigid rotation of 15° about a C_{sp³}-C_{sp³} bond and by ~ 5 kcal/mol for a 60° rotation. Before applying these restraints, however, we first optimized the hydrogen atom positions to ensure that no restraints involved unrealistic torsion angles.

The resultant comparisons yielded rms deviations of 0.023 Å for 40,562 bond lengths and of 1.96° for 58,087 bond angles. These deviations are respectable, though smaller rms deviations of 0.014 Å and 1.2° were found in comparing MMFF94 to experiment for 30 small molecules³ and still smaller rms deviations of 0.006 Å and 1.16° were found in

comparisons to MP2/6-31G*-optimized geometries.¹ Especially for bond lengths, the larger deviations may partly reflect cases in which MMFF94 is unable to fully represent a real chemical variation present in the CSD data. One such case arose, for example, for aliphatic C—F bonds, where lengths of about 1.32 Å in trifluoromethyl groups rise sharply to ~ 1.40 Å when the carbon atom is singly fluorinated. As previously discussed,³ MMFF94 gives an intermediate value as a result of an intentionally chosen compromise. Other, similar cases undoubtedly arise.

Not all of the discrepancy between the MMFF94 and CSD geometries necessarily represents deficiencies in the MMFF94 model. Part can be attributed to the use of averaged atomic positions in defining X-ray values for bond lengths and angles. In this connection, Clark et al.⁴¹ assert that X-ray bond lengths can be too short by 0.005–0.015 Å if the structure is not corrected for rigid body motions. In agreement with their assessment, we find that the MMFF94 bond lengths (most of which are principally determined by MP2/6-31G*-based reference values) exceed the CSD bond lengths by 0.010 Å on average. Second, part of the larger apparent errors encountered in reproducing the CSD geometries may be attributed to uncertainties in the experimental values themselves. There appears to be substantial variation in the quality of crystallographic structures, even when determined at low *R* factors.^{41,42}

FIT TO CSD DATA PREVIOUSLY USED IN ASSESSING MOLECULAR FORCE FIELDS

For use in characterizing the Tripos 5.2 force field, Clark et al.⁴¹ a few years ago extracted from the then-current version of the CSD the first 76 structures having *R* factors below 0.05. More recently, Mayo et al. used the same structures to benchmark their DREIDING force field,³² and Momany and Rone⁴³ employed 49 of them to help assess their QUANTA 3.2 parameterization for CHARMM. To test MMFF94, we extracted the same structures and prepared them in the manner previously described. Interestingly, only 5 structures appear in the larger set of 2861 structures used in extending MMFF94's parameterization. As had other investigators, we removed counterions and coordinating water molecules; when two chemically identical molecules occupied the unit cell, we arbitrarily deleted one. Like the groups just cited, we report deviations only for bonds, angles, torsions, and atomic positions that do not involve

hydrogen atoms, and we compare MMFF94 to CSD atomic positions after rigid least-squares superposition of the heavy atoms. In their studies, the Tripos group used a conjugate-gradient algorithm and refined each structure until a rms gradient over all atoms of 0.1 kcal/mol · Å was reached.⁴¹ Mayo et al. employed the same convergence criterion but used a variable-metric algorithm.⁴⁴ Momany and Rone, in contrast, employed a powerful Newton–Raphson procedure and used a much lower rms gradient of 1×10^{-6} kcal/mol · Å as the convergence criterion. Our optimizations also used a variable-metric algorithm¹ but were carried to a rms gradient of 1×10^{-6} kcal/mol · Å. Thus, the MMFF94 structures, like those of Momany and Rone, rigorously represent stationary points (and almost certainly are energy minima).

Detailed MMFF94 results are presented in Table XIII for 75 of the 76 CSD structures; one, the structure with the CSD refcode of "AFMSCY," was a sulfonium salt (R_3S^+) that could not be handled by MMFF. As noted earlier in this study, we had decided at the outset not to include such species in the extended parameterization, though it would not have been difficult to do so. Chemical names for the CSD structures may be found in the article by Clark and coworkers⁴¹; the structures are depicted there and by Mayo et al.³² The numbers of atoms, bonds, angles, and torsions in each structure are the same as those listed by Mayo et al.

In examining Table XIII, we note, first of all, that the rms deviation of 0.021 Å for bond lengths is virtually the same as that found in the previous subsection for the larger set of CSD structures, as is the rms deviation of 1.97° for bond angles. These rms deviations are about 60% as large as those of 0.035 Å and 3.224° reported for the DREIDING force field.³² Thus, the more specifically parameterized MMFF94 force field indeed describes these aspects of the molecular geometry more accurately. For comparison, Clark et al. report rms deviations of 0.025 Å for bond lengths and 2.50° for bond angles for the Tripos 5.2 force field; Momany and Rone do not report results for bond and angle deviations. We consider the MMFF94 deviations to be quite respectable and note that perfect agreement with the CSD geometries would not be expected.

When deviations in torsion angles and in atomic positions are considered, a different, and more complex, picture emerges. Thus, MMFF94's rms deviations are 16.46° and 0.474 Å, respectively, while those reported for DREIDING are 8.948° and 0.235 Å, only about half as large. Similarly, Clark

TABLE XIII.
Comparison of MMFF94 and Crystallographic Structures^a

System	Natm	Ratm	Matm	Rbnd	Abnd	Mbnd	Rang	Mang	Rtor	Mtor
Total	75 ^b	0.474	3.657	0.021	0.009	-0.106	1.97	-11.77	16.46	159.25
AAXTHP	26	1.317	3.657	0.024	0.013	0.045	2.79	-6.89	53.29	131.57
ABAXES	25	0.156	0.389	0.021	0.003	-0.043	1.57	-3.59	4.17	-12.88
ABBUMO10	22	0.053	0.097	0.012	0.008	0.024	1.07	2.21	2.12	-5.65
ABINOR02	10	0.062	0.117	0.015	0.006	0.030	1.89	-3.65	4.68	8.15
ABINOS01	10	0.070	0.111	0.014	0.012	0.025	2.17	4.31	3.69	5.61
ABTOET	27	0.400	0.935	0.020	0.009	0.046	1.95	4.12	14.41	-50.86
ABZTCX	20	0.766	2.242	0.022	0.008	0.075	3.30	-11.77	36.65	124.31
ACADOS	22	0.160	0.325	0.017	0.008	0.052	1.13	3.54	5.01	-16.13
ACAFRL	17	0.305	0.921	0.024	0.003	-0.081	1.73	4.07	10.20	-43.88
ACANIL01	10	0.127	0.231	0.011	0.005	0.028	1.60	4.05	7.29	-16.99
ACARAP	22	0.561	1.708	0.021	0.018	0.037	1.87	3.85	16.49	59.09
ACBNZA01	13	0.194	0.431	0.025	0.015	0.042	2.44	5.29	10.04	20.34
ACBUOL	24	1.173	2.706	0.017	0.009	0.042	2.73	-8.03	55.26	139.10
ACCITR10	22	0.143	0.537	0.019	0.006	-0.037	1.71	3.90	5.91	-20.87
ACDXUR	19	0.291	0.860	0.021	0.001	-0.049	1.69	-4.10	8.13	19.97
ACENAP03	12	0.027	0.049	0.024	-0.007	-0.048	1.34	-3.05	0.48	0.99
ACFPCH	17	0.475	0.910	0.019	0.011	0.040	1.40	-3.46	21.60	52.88
ACFUCN	14	0.163	0.261	0.027	0.016	0.058	2.03	-4.88	7.85	-12.78
ACGLSP	25	0.475	1.350	0.027	0.022	0.052	1.75	3.12	12.99	41.23
ACGLUA11	15	0.221	0.630	0.016	0.008	0.035	2.11	4.28	7.42	22.16
ACHGAL	17	0.236	0.736	0.017	0.011	0.033	1.70	4.48	6.94	-23.69
ACHIST20	14	1.205	2.035	0.030	0.013	0.095	3.03	8.65	41.76	-88.77
ACHNAP10	15	0.060	0.094	0.024	0.021	0.040	1.82	5.50	2.10	4.51
ACHTAR10	10	0.197	0.369	0.026	0.023	0.040	1.33	-3.30	7.97	19.87
ACIMDC	4	0.018	0.024	0.007	0.006	0.010	1.82	2.55	—	—
ACINDN	14	0.050	0.082	0.021	0.009	0.039	1.54	2.78	1.84	4.84
ACINST	26	0.352	1.070	0.022	0.019	0.042	1.88	-5.87	9.34	31.01
ACKYNU	18	0.487	1.176	0.025	0.012	0.066	1.94	-4.23	15.24	28.83
ACMBPN	24	0.230	0.704	0.021	0.015	0.048	1.98	4.71	11.41	40.58
ACMEBZ	14	0.217	0.506	0.024	0.015	0.055	2.34	-5.04	8.73	17.90
ACMTDE	16	0.158	0.255	0.021	0.010	0.054	1.00	-3.08	6.51	-10.83
ACNORT	31	0.195	0.440	0.018	0.011	0.034	2.18	-5.43	5.85	-13.24
ACNPAC10	15	0.057	0.144	0.025	0.002	0.049	1.87	5.42	1.76	5.06
ACNPEC	21	0.832	2.163	0.027	0.022	0.055	2.20	-4.84	28.74	68.64
ACONTN10	32	0.175	0.783	0.017	0.004	0.057	1.42	4.01	3.98	21.17
ACPENC10	15	0.449	1.136	0.022	0.003	-0.045	2.38	-7.48	12.18	-40.88
ACPPCA	12	0.178	0.396	0.027	0.020	0.057	2.17	-4.36	10.04	-18.04
ACPRET03	29	0.292	0.695	0.021	0.009	0.051	1.31	-4.44	5.25	-12.53
ACPYNS	18	0.370	0.603	0.024	0.017	0.054	1.53	-4.06	12.51	21.72
ACRAMS	26	1.071	1.810	0.028	0.010	0.077	1.72	-5.42	16.51	-72.42
ACSALA01	13	0.075	0.171	0.024	0.015	0.055	2.42	5.61	2.49	5.39
ACSESO10	23	0.171	0.526	0.019	0.012	0.037	1.25	2.93	8.22	20.53
ACTAND	24	0.174	0.502	0.015	0.010	0.034	1.18	-3.30	3.33	15.96
ACTHBZ	22	0.560	1.662	0.015	0.007	0.039	1.74	4.60	18.58	57.72
ACTHCP	10	0.085	0.158	0.019	-0.006	-0.037	1.56	3.27	6.52	11.21
ACTOLD	11	0.140	0.299	0.019	0.014	0.042	1.50	3.77	7.33	19.94
ACTYSN	16	0.381	0.895	0.018	0.005	0.048	1.88	-5.07	13.97	-32.99
ACURID	19	0.414	0.792	0.014	0.000	-0.032	1.89	-4.26	14.00	-34.30
ACVCHO	12	0.048	0.096	0.015	0.009	0.033	1.29	-3.05	1.92	3.29
ACXMOL	22	0.532	1.255	0.021	0.010	0.050	2.32	5.79	13.05	-44.33
ACXMPR	14	0.972	2.609	0.020	0.008	0.035	3.72	-9.93	66.61	159.25

(Continues on next page)

et al. report rms deviations of 9.65° for torsion angles and of 0.25 \AA for atomic positions for the Tripos 5.2 force field, again in much better agreement with the CSD data. On the other hand, Momany and Rone found a rms deviation in atomic positions of 0.438 \AA for their CHARMM parameterization, only slightly better than that for MMFF94; they did not report torsion angles. In assessing such differences in apparent accuracy, Momany and Rone pointed out that crystal geometries often display torsion angles that differ markedly from those that would be found for an isolated molecule in vacuum, the conditions under which the force-field optimizations were carried out. They further noted that the Tripos 5.2 and DREIDING studies omitted electrostatic interactions, precisely the ones most likely to induce a change in conformation when a structure is removed from geometrically confining and/or elec-

trostatically stabilizing influences in the crystal lattice. They also suggested that the relatively weak convergence criteria used by Clark et al. and by Mayo et al. might have yielded partially refined structures in which the torsion angles had not yet approached their equilibrium values.

To test their hypotheses, Momany and Rone reoptimized 14 of the most troublesome CSD structures in the context of the crystallographic environment. They found that the rms deviation in atomic positions for these 14 structures fell by half, from 0.802 to 0.399 \AA . Moreover, they found a rms deviation of just 0.376 \AA for the same set of structures when they omitted electrostatic interactions and refined the structures in vacuum to a gradient of $0.1 \text{ kcal/mol} \cdot \text{\AA}$ rms. In a similar fashion, we find that MMFF94's rms deviation for torsion angles falls from 16.46° to 13.64° when the convergence cutoff is raised to $0.1 \text{ kcal/mol} \cdot \text{\AA}$, and then

TABLE XIII.
(continued)

System	Natm	Ratm	Matm	Rbnd	Abnd	Mbnd	Rang	Mang	Rtor	Mtor
ACYGLY11	8	0.071	0.119	0.027	0.019	0.044	2.33	-3.69	2.54	-3.53
ACYTID	17	0.171	0.304	0.025	0.010	0.057	1.79	4.03	4.97	13.68
ADELOX10	28	0.168	0.507	0.014	0.009	-0.034	1.56	-4.82	5.27	21.92
ADENOS10	19	0.201	0.517	0.018	0.007	0.057	1.29	-2.85	6.03	-15.51
ADFGLP	11	0.679	1.454	0.024	0.003	0.004	1.82	3.75	43.27	73.72
ADGSMH	29	0.440	1.158	0.021	0.017	0.043	2.05	4.36	11.97	40.33
ADHELA10	14	0.098	0.200	0.013	0.003	-0.022	2.40	4.24	5.23	-9.96
ADMANN	12	0.170	0.339	0.017	0.014	0.028	2.47	4.88	6.57	10.91
ADMHEP	14	0.083	0.162	0.012	0.004	0.027	1.97	3.82	3.97	-9.94
ADMINA	16	0.054	0.116	0.022	0.002	0.042	1.66	-3.84	1.98	6.09
ADMOPM	24	0.989	2.148	0.028	0.012	0.078	3.09	-6.98	29.22	112.49
ADRTAR	13	0.747	1.601	0.015	0.004	0.026	1.72	3.30	38.92	86.95
ADYPNL	33	0.499	1.087	0.014	0.009	-0.026	1.36	-3.84	13.45	-46.82
AEBDOD10	22	0.173	0.501	0.010	0.003	-0.023	1.28	-3.63	7.47	16.98
AENLAN10	35	0.531	1.898	0.032	0.008	-0.106	2.07	10.22	7.97	-42.08
AFCYDP	24	0.319	0.686	0.041	0.009	0.105	2.86	7.73	8.18	-20.18
AFURPO10	13	0.095	0.182	0.016	0.004	-0.024	3.24	-8.12	5.52	-13.96
AFUTDZ10	12	0.052	0.090	0.027	0.013	0.055	2.30	-4.52	3.59	7.76
AFUTHU	16	0.353	0.655	0.020	0.002	-0.038	2.17	-4.36	12.35	-29.10
AGALAM10	15	0.216	0.647	0.023	0.016	0.057	1.74	-3.14	6.85	-22.23
AGLUAM10	13	0.334	0.746	0.024	0.013	0.048	2.11	-4.11	12.88	-29.10
AHARFU	16	0.215	0.583	0.020	-0.005	-0.035	2.15	6.49	7.15	17.04
AHCDLA	14	0.049	0.078	0.019	0.003	-0.051	1.66	-3.93	2.08	-4.98
AHDITX	26	0.146	0.471	0.014	0.004	0.026	1.61	-6.47	3.17	13.46

^aAtomic displacements and bond deviations in angstroms, angle and torsion deviations in degrees. The listed quantities give the number (N) of atoms in each molecule, followed by the rms deviation (R), average signed deviation (A), and maximum deviation (M) in atomic positions (atm), bond lengths (bnd), bond angles (ang), and torsion angles (tor); the "atm" quantities measure displacements in atomic positions. The notation corresponds to that used in ref. 32.

^bNumber of molecules. The total number of bonds, angles, and torsions are 1461, 2140, and 2383, respectively; the numbers for each molecule are the same as those listed in ref. 32.

to 10.03° when electrostatic interactions are also suppressed. Concomitantly, the rms deviation in atomic positions falls first from 0.474 to 0.424 Å and then to 0.336 Å. While these rms deviations remain larger than those reported for the Tripos 5.2 and DREIDING force fields, the differences are narrowed very considerably. Moreover, the simple device of omitting electrostatic interactions can be expected to unbalance a force field like CHARMM or MMFF94 that relies on a carefully orchestrated interplay between electrostatic, torsional, and vdW contributions. Finally, we note that Mayo et al. also reported comparisons in which electrostatic interactions based on Gasteiger charges were used. These calculations did show larger rms deviations in atomic positions, but only modestly so. Although Gasteiger charges are much smaller than MMFF94 charges,² this last observation is not so readily explained. However, the effect of the partial optimization to a gradient of 0.1 kcal/mol·Å may carry more weight than we have credited to it, as the properties of the optimizer itself come into play whenever the optimization fails to reach a stationary point on the potential energy surface. In this regard, we note that the Newton–Raphson optimizer used by Momany and Rone possesses complete information about the curvature of the energy surface and immediately generates search directions that include complicated (in Cartesian space) torsional motions. In comparison, the optimizer used in OPTIMOL/MMFF, like that used by Mayo et al., employs a variable-metric algorithm. However, our algorithm, which we believe to be highly efficient, employs a block-diagonal estimate for the inverse Hessian matrix, which is then successively updated via the BFGS scheme.⁴⁵ It may yield a more nearly complete torsional optimization at a rms gradient of 0.1 kcal/mol·Å than does the DREIDING optimizer, which we understand uses a simple diagonal guess for the initial inverse Hessian.

Admittedly, the foregoing is conjecture: What we can say with certainty is that a comparison of torsion angles and atomic positions is fraught with difficulties. We feel that such comparisons have little meaning unless (i) the conditions of the experimental measurements are properly emulated (e.g., through explicit crystal-phase simulations), and if geometry optimization is used, unless (ii) the minima on the potential energy surface are accurately located.

Concluding Discussion

This article describes the extension of MMFF94 to a much broader range of pharmaceutical and chemical structures than was covered by the core parameterization presented in parts II,² III,³ and IV.⁴ One key to this extension was the extraction and use of approximately 2800 good-quality crystallographic structures from the Cambridge Structural Database. We also generated HF/6-31G*-optimized geometries for a variety of small molecules and ions, and obtained MP2/6-31G*-optimized geometries for these (Table II) and for a number of additional structures (Tables IV and VIII). The parameters derived from this additional computational data allow charge distributions, geometries, and conformational energies of sulfonamides and certain oxyphosphorous compounds to be described at a high level. For prospective pharmaceutical and bio-organic applications, compounds of these types are among the most important of those not included in MMFF94's core parameterization.

A series of carefully calibrated empirical rules and default-parameter assignments plays a crucial role in MMFF94's extension. Given the importance of electrostatic contributions in modeling intermolecular interactions, key among these is an empirical rule used for assigning otherwise unknown bond charge increments from "partial bond charge increments" [eq. (17)]. These quantities, which essentially reflect effective electronegativities of the MMFF atom types, were determined by requiring that they reproduce the explicitly parameterized bond charge increments as well as possible.

For bonds and angles, the extension of MMFF94 proved to be quite straightforward. We determined additional reference bond lengths and angles by first fitting to the geometries of the MP2/6-31G*-optimized structures, and by then fitting to the geometries of the CSD structures. We then developed empirical rules for reference bond lengths [eq. (18)] and for bond-stretching force constants [eq. (19)], and specifically calibrated each to reproduce core MMFF94 parameters. For reference bond angles, we generated a default angle for each central-atom type by taking suitable averages over the parameterized reference angles. We then employed a previously published empirical rule²⁹ to obtain the angle-bending force constant [cf. eq.

(20)]. This empirical rule had previously been calibrated to reproduce spectroscopically observed force constants in a wide variety of molecules.

Also described are procedures used to extend MMFF94's parameterization for stretch-bend interactions, out-of-plane bending interactions, torsion interactions, and van der Waals interactions. For torsion interactions, we began by fitting additional torsion parameters to MP2/6-31G*-derived conformational energies for sulfonamides and oxyphosphorous compounds, among others. To further extend the parameterization, we first generated default values that depend only on the atom types for the two central atoms (and on the "torsion type" index) by suitably averaging over the torsion parameters obtained in explicit fits to *ab initio* conformational energies. We then formulated an empirical rule for generating additional default ("wildcarded") parameters. This rule is patterned after ones previously devised for the DREIDING³² and UFF³³ force fields but differs in that it employs slightly different functional forms, makes finer context-related distinctions, and has been calibrated specifically for MMFF94 by requiring that the generated torsion parameters closely reproduce those obtained in the explicit parameterizations.

The extension of the van der Waals parameterization also included the additional atom types needed to accommodate ionic species such as "protein" metals Zn^{2+} , Ca^{2+} , Cu^{2+} , Cu^+ , Fe^{2+} , and Fe^{3+} ; the hydronium and hydroxide ions; and various halide, alkali, and alkali earth ions. Only a preliminary parameterization was undertaken but MMFF94 results for hydrates (Tables XI and XII) appear to obey reasonable trends in interaction distances and, where experimental data were utilized, in association energies. It is well to keep in mind, however, that "electrostatic" interactions such as those between ligands and protein metals (e.g., Zn^{2+}) may need to be treated by a model that takes account of partial chemical bond formation and of concomitant charge transfer between the coordinating sites.

Assessments of performance of the extended MMFF94 parameterization using the full set of CSD data and using a smaller set of CSD structures previously employed for force-field evaluation showed that MMFF94 accounts for crystallographically derived bond lengths and bond angles with rms deviations of ~ 0.02 Å and $\sim 2^\circ$. We also compared deviations in torsion angles and atomic positions for the smaller set of CSD data.

Here, not surprisingly,⁴⁶ we found that MMFF94's *in vacuo* optimization often changed the conformation considerably from that observed in the crystal. It would be of interest to determine how well MMFF94 reproduces crystal geometries in energy minimization studies—or, better, in molecular-dynamics simulations—when the calculations are carried out in the context of the crystal lattice.

To date, MMFF has principally been used in energy-minimization studies and in other calculations and analyses supported by the host program OPTIMOL.^{1,47} Recently, however, a version of MMFF has been implemented in the commercially distributed program CHARMM,^{43,48} and MMFF94 is expected to become available in CHARMM,⁴⁹ the academic version, as well. In addition, work is currently underway to implement MMFF94 in the BatchMin module of the MacroModel program suite.⁵⁰ Given the good results obtained for conformational energies, including results for conformational comparisons pertinent to protein conformation,⁴ and given the careful attention devoted to achieving a balanced representation of intermolecular interactions,² we consider MMFF94 to be both a "protein" force field and an "organic" force field. We expect it to do well in modeling structures and properties in both arenas.

Acknowledgments

I thank Joseph Andose for implementing an earlier version of eq. (18), and Robert Nachbar and Simon Kearsley for providing procedures that assisted in the preparation of the CSD data and in rigidly superposing molecular structures.

Supplementary Material

Computer-readable parameter files corresponding to Tables I, III, V, and VI are available in Supplementary Material as Appendix B to part I of this study.¹⁵

References

1. Part I: T. A. Halgren, *J. Comput. Chem.* (this issue).
2. Part II: T. A. Halgren, *J. Comput. Chem.* (this issue).
3. Part III: T. A. Halgren, *J. Comput. Chem.* (this issue).
4. Part IV: T. A. Halgren and R. B. Nachbar, *J. Comput. Chem.* (this issue).

5. The Cambridge Crystallographic Database is obtainable from the Cambridge Crystallographic Data Centre, Cambridge, UK.
6. R. B. Nachbar, Merck Research Laboratories (unpublished research).
7. AMF is the proprietary Advanced Modeling Facility developed at the Merck Research Laboratories by J. D. Andose, R. A. Blevins, E. M. Fluder, J. Shpungin, and others in the Scientific Computing Group of the Department of Molecular Design and Diversity.
8. T. A. Halgren, *J. Am. Chem. Soc.*, **114**, 7827–7843 (1992).
9. M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian 92 (Revision C), Gaussian, Inc., Pittsburgh, PA, 1992.
10. C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, **11**, 361 (1987).
11. PROBE is a computer program used to derive molecular-mechanics parameters in least-squares fits to data obtained from *ab initio* calculations. PROBE was created for the Biosym Consortium on Potential Energy Functions by Biosym Technologies, Inc. (San Diego, CA); cf. J. R. Maple, M.-J. Hwang, T. P. Stockfish, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler, *J. Comput. Chem.*, **15**, 161–182 (1994).
12. In addition to the 27 values constrained in the PROBE fit to CHELPG-based values, the “#X94” bond charge increments also include parameters for homopolar bonds (having zero bond charge increments) and parameters for bonds to bromine and iodine from carbons of MMFF atom types 1 and 2. Owing to the unavailability of 6-31G* basis sets for Br and I in Gaussian 92, we estimated these quantities from scaled values (increased by 20%) of observed dipole moments for the methyl and vinyl halides.
13. R. J. Abraham, L. Griffiths, and P. Loftus, *J. Comput. Chem.*, **3**, 407–416 (1982); R. J. Abraham, G. H. Grant, I. S. Haworth, and P. E. Smith, *J. Comput. Aid. Mol. Design*, **5**, 21–39 (1991).
14. L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, NY, 1960, Chapter 3.
15. The MMFF94 parameters are included in Appendix B (Supplementary Material) for Part I (ref. 1) and are available in computer-readable ASCII format.
16. V. Schomaker and D. P. Stevenson, *J. Am. Chem. Soc.*, **63**, 37–40 (1941).
17. R. Blom and A. Haaland, *J. Mol. Struct.*, **128**, 21–27 (1985).
18. Eq. (18) was developed by Joseph D. Andose of the Merck Research Laboratories for use with the predecessor MM2X force field (cf. ref. 1). The parameters and the adjustments for hybridization and bond-order described in the subsection “Empirical Rule For Reference Bond Lengths” include modifications to parameters and procedures he developed.
19. A. L. Allred and E. g. Rochow, *J. Inorg. Nucl. Chem.*, **5**, 264 (1958).
20. *Table of Periodic Properties of the Elements*, E. H. Sargent & Co., Brochure S-18806, Dyna-Slide Co., © 1962.
21. Compare with ref. 14, Table 7-2, p. 224.
22. See, for example, J. Stals, *Rev. Pure Appl. Chem.*, **21**, 1–22 (1970).
23. R. M. Badger, *J. Chem. Phys.*, **2**, 128 (1934); R. M. Badger, *J. Chem. Phys.*, **3**, 710 (1935).
24. R. S. Roy, *Proc. Phys. Soc., Ser. 2*, **1**, 445–448 (1968).
25. D. M. Byler, H. Susi, and W. C. Damert, *Spectrochim. Acta A*, **43**, 861–863 (1987).
26. M. R. Peterson and I. G. Csizmadia, *J. Mol. Struct.*, **123**, 399–412 (1985); cf. Table I of this paper.
27. Badger’s rule (cf. ref. 28) also performed reasonably well, but gave a somewhat higher rms deviation of 1.06 md/Å.
28. D. R. Herschbach and V. W. Laurie, *J. Chem. Phys.*, **35**, 458–463 (1961).
29. T. A. Halgren, *J. Am. Chem. Soc.*, **112**, 4710–4723 (1990).
30. These calculations used the MM2X force field and the sulfonamide-based supplementary parameter file “SULF.PAR” described in: M. K. Holloway, J. M. Wai, T. A. Halgren, P. M. D. Fitzgerald, J. P. Vacca, B. D. Dorsey, R. B. Levin, W. J. Thompson, L. J. Chen, S. J. deSolms, N. Gaffin, A. K. Ghosh, E. A. Giuliani, S. L. Graham, J. P. Guare, R. W. Hungate, T. A. Lyle, W. M. Sanders, T. J. Tucker, M. Wiggins, C. M. Wiscount, O. W. Woltersdorf, S. D. Young, P. L. Darke, and J. A. Zugay, *J. Med. Chem.*, **38**, 305–317 (1995).
31. In a few cases, information on crystal-structure geometries was used to modify the empirical-rule values for delocalized single bonds. Such torsion parameters are identified by the label “#E94.”
32. S. L. Mayo, B. D. Olafson, and W. A. Goddard III, *J. Phys. Chem.*, **94**, 8897 (1990).
33. A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, *J. Am. Chem. Soc.*, **114**, 10024–10035 (1992).
34. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535–542 (1977); E. E. Abola, F. C. Bernstein, S. H. Bryant, T. F. Koetzle, and J. Weng, In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications*, F. H. Allen, G. Bergeroff, and R. Sievers, Eds.; Data Commission of the International Union of Crystallography: Cambridge, UK, 1987, pp. 107–132.
35. P. Kebarle, *Annu. Rev. Phys. Chem.*, **28**, 445 (1977), and references therein.
36. L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, NY, 1960, Chapter 13.
37. M. C. Day Jr. and J. Selbin, *Theoretical Inorganic Chemistry*, 2nd ed., Van Nostrand Reinhold, New York, 1969, p. 120.
38. J. Chandrasekhar, D. C. Spellmayer, and W. L. Jorgensen, *J. Am. Chem. Soc.*, **106**, 903–910 (1984).
39. The same assumption was previously employed in the core MMFF94 parameterization of polar intermolecular interactions (ref. 2). It presumes that additional polarization in the condensed phase should strengthen favorable electrostatic interactions.
40. See, for example: R. F. Frey, J. Coffin, S. Q. Newton, M. Ramek, V. K. W. Cheng, F. A. Momany, and L. Schafer, *J. Am. Chem. Soc.*, **114**, 5369–5377 (1992).
41. M. Clark, R. D. Cramer III, and N. Van Opdenbosch, *J. Comput. Chem.*, **10**, 982 (1989).
42. P. G. Jones, *Chem. Soc. Rev.*, **13**, 157 (1984).
43. F. A. Momany and R. Rone, *J. Comput. Chem.*, **13**, 888–900 (1992).
44. B. D. Olafson, personal communication.

45. R. Fletcher, *Practical Methods of Optimization: Unconstrained Optimization*, Vol. 1, Wiley, New York, 1980, Chapter 3.
46. M. C. Nicklaus, S. Wang, J. S. Driscoll, and G. W. A. Milne, *Biorganic & Medicinal Chemistry*, **3**, 411–428 (1995).
47. Current intentions are to release OPTIMOL/MMFF94 through the Quantum Chemistry Program Exchange (Indiana University) under terms that would permit its unrestricted usage but prohibit its commercialization.
48. R. Czerminski, T. A. Halgren, and J. L. Banks, *unpublished research*. Currently, a version CHARMM that includes the earlier and less widely parameterized MMFF93 force field (refs. 3, 4; see also "Extension of the Parameterization for van der Waals Interactions" section of this study) is available from Molecular Simulations, Inc. (San Diego, CA). However, while the local Merck code employs MMFF94, arrangements for including MMFF94 in the distributed MSI version have not yet been concluded.
49. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187–217 (1983).
50. P. S. Shenkin and T. A. Halgren, work in progress. The MacroModel program suite, developed in the laboratories of Professor Clark Still, is available from Columbia University.